

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Efficient quantum-chemical
methods for calculating NMR
shieldings in second-order
Møller-Plesset
perturbation theory**

Marina Maurer

aus

Karlsruhe

2014

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. C. Ochsenfeld betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 4. August 2014

(Marina Maurer)

Dissertation eingereicht am: 07.08.2014

1. Gutachter: Prof. Dr. Christian Ochsenfeld

2. Gutachter: Prof. Dr. Hubert Ebert

Mündliche Prüfung am: 17.10.2014

Danksagung

Bei Herrn Prof. Dr. Christian Ochsenfeld bedanke ich mich herzlich für die Bereitstellung des interessanten Themas und die beständige Unterstützung während meiner Dissertation in seinem Arbeitskreis.

Herrn Prof. Dr. Hubert Ebert danke ich für sein freundliches Entgegenkommen, das Zweitgutachten anzufertigen.

Bei den Mitgliedern des Arbeitskreises bedanke ich mich für die freundliche Arbeitsatmosphäre, die vielfältige Hilfsbereitschaft und Zusammenarbeit. Im Besonderen möchte ich mich bei Dr. Matthias Beer, Dr. Denis Flaig, Dr. Simon Maurer, Dr. Keyarash Sadeghian und Dr. Jörg Kussmann bedanken.

Meiner Familie, meinem Freund und meinem Freundeskreis danke ich für die fortwährende Hilfe und den stetigen Rückhalt.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Entwicklung von effizienten quantenchemischen Methoden zur Berechnung von Eigenschaften großer molekularer Systeme. Der Schwerpunkt der Arbeit liegt dabei auf der Optimierung von Methoden zur Berechnung von NMR Abschirmungen. Diese können mit quantenchemischen Methoden berechnet werden, in dem man die elektronische Energie nach dem Magnetfeld und dem kernmagnetischen Moment ableitet. Um NMR Abschirmungen von großen Molekülen in einer praktikablen Zeit berechnen zu können, muss der Rechenaufwand so klein wie möglich gehalten werden. Leider steigt der Rechenaufwand von konventionellen quantenchemischen Methoden extrem mit der Molekülgröße an, was die Berechnungen mit diesen Methoden auf kleine Moleküle beschränkt. Im Vergleich dazu, wird in der vorliegenden Arbeit eine Methode vorgestellt, mit der der Rechenaufwand für die Berechnung der NMR Abschirmung eines Kerns aus einem ausreichend großen Systems konstant bleibt, wenn man das Molekül noch weiter vergrößert. Dieses so genannte sublineare Skalenverhalten des Rechenaufwandes mit der Molekülgröße lässt sich verwirklichen in dem man das lokale Verhalten der kernmagnetischen Störung in den zeitbestimmenden Schritten der Berechnung ausnutzt. Um dieses lokale Verhalten effizient ausnutzen zu können, basiert die Methode auf der lokalen Atomorbital-basierten Möller-Plesset Störungstheorie in zweiter Ordnung (AO-MP2), die auch Teile der Elektronenkorrelationseffekte erfasst, die in der Hartree-Fock (HF) Methode vernachlässigt werden.

Desweiteren müssen Abschätzungen für Zweielektronenintegrale angewendet werden um den sublinear skalierenden Rechenaufwand für die AO-MP2 NMR Abschirmungen erreichen zu können. Zweielektronenintegrale beschreiben im Allgemeinen die Wechselwirkung zwischen zwei Ladungsverteilungen und werden fast immer in den aufwändigsten Schritten einer Berechnung gebraucht. Deswegen ist es wichtig, signifikante Beiträge abschätzen zu können, um daraufhin nur diese berechnen zu müssen. Für die Abschätzung dieser Beiträge werden hier QQR-Abschätzungen verwendet, welche ursprünglich für Energieberechnungen entwickelt wurden. In dieser Arbeit werden die QQR-Abschätzungen für AO-MP2 NMR Abschirmungen erweitert. Indem man diese Abschätzungen anwendet, lässt sich der Rechenaufwand der zeitbestimmenden Schritte der AO-MP2 NMR Berechnun-

gen auf ein sublineares Skalenverhalten für große Systeme reduzieren. Desweiteren wurden die QQR-Abschätzungen für die Berechnung von AO-MP2 Gradienten erweitert, welche für die Optimierung von Strukturen von Molekülen verwendet werden können. Die Anwendung dieser Abschätzungen in der Berechnung von AO-MP2 Gradienten führt zu einem asymptotischen linear skalierendem Rechenaufwand mit der Molekülgröße.

Obwohl die MP2 basierten NMR Abschirmungen meistens schon recht genaue Ergebnisse liefern, lässt sich die Genauigkeit der Ergebnisse noch weiter verbessern. Für Energien konnte gezeigt werden, dass die Genauigkeit der MP2 Methode sich verbessern lässt indem man die Beiträge entgegengesetzten und gleichen Spins der MP2 Energie unterschiedlich skaliert. Ähnlich zu diesem Ansatz, werden Methoden in dieser Arbeit vorgestellt bei denen zum ersten Mal die unterschiedlichen Spin Beiträge von NMR Abschirmungen skaliert werden. Die neuen skalierten MP2 NMR Verschiebungen zeigen in den meisten Fällen eine Verbesserung der Genauigkeit im Vergleich zu unskalierten MP2 Verschiebungen, was für Kohlenstoff, Phosphor, Fluor, Stickstoff und Sauerstoff NMR Verschiebungen getestet wurde.

Abstract

In this thesis, contributions to efficient quantum-chemical methods for calculating properties of large molecular systems are presented. The main focus of this work is on the optimization of methods for calculating NMR shieldings. These can be computed with quantum-chemical methods by differentiation of the electronic energy with respect to the magnetic field and the nuclear magnetic spin moment. To calculate NMR shieldings of larger molecules in a feasible time, the computational cost of the calculation must be kept as small as possible. Unfortunately, the computational cost of conventional quantum-chemical methods increases rapidly with the molecular size, which limits the calculations with these methods to small molecules. In comparison, a method is presented within this work by which the computational cost of calculating the NMR shielding of a nucleus in a sufficiently large molecule stays constant once a certain molecule size has been reached. This so-called sublinear-scaling behavior with molecular size becomes accessible by exploiting the local behavior of the nuclear magnetic spin perturbation in the rate-determining steps of the calculation. To exploit this local behavior efficiently, the method is based on the local atomic orbital-based second-order Møller-Plesset perturbation theory (AO-MP2), which includes also parts of electron correlation effects missing in the Hartree-Fock (HF) method.

Furthermore, estimates for two-electron integrals need to be applied to achieve the sublinear-scaling computational cost for the AO-MP2 NMR shieldings. In general, these integrals describe the interactions between two charge distributions and they are almost always needed in the most time-consuming steps of a calculation. Therefore, it is important to estimate and only compute the significant contributions of the two-electron integrals. To this extent, the QQR-type integral estimates are exploited, which were originally developed for energy calculations. In this work, they are extended to AO-MP2 NMR shieldings. By applying these estimates, the computational cost of the rate-determining steps of the AO-MP2 NMR calculations can be reduced to sublinear for large molecules. Moreover, the QQR-type integral estimates are also extended for calculating AO-MP2 gradients, which can be used to optimize the structure of molecules. The application of these estimates in the calculations of AO-MP2 gradients leads to an asymptotic linear-scaling behavior with the system size.

While MP2 NMR shieldings are most of the times quite accurate, their accuracy can be improved even further. For energies, it has been shown that the accuracy of the MP2 method can be improved by scaling the opposite and same spin parts of the MP2 energy differently. Similar to this approach, methods are presented in this work, which scale for the first time the opposite and same spin contributions of MP2 NMR shieldings. The new scaled MP2 NMR shifts show in most cases an improvement of the accuracy in comparison to nonscaled MP2 NMR shifts, which is shown for carbon, phosphorus, nitrogen, oxygen, and fluorine NMR shifts.

Contents

List of publications	1
1 Introduction	3
2 Theory	7
2.1 Møller-Plesset perturbation theory	7
2.1.1 Molecular orbital-based MP2 energy expression	7
2.1.2 Atomic orbital-based MP2 energy expression	8
2.1.3 Integral estimates	9
2.2 AO-based MP2 energy gradients	11
2.2.1 The first derivative with respect to the nuclear coordinates ξ	11
2.2.2 Avoiding the computation of the perturbed pseudo-densities	12
2.2.3 AO-based Z-vector method for avoiding the first-order density matrix	15
2.2.4 Integral estimates for AO-MP2 gradients	16
2.3 AO-based MP2 NMR shieldings	18
2.3.1 The AO-MP2-based magnetic shielding tensor	19
2.3.2 AO-based Z-vector method for avoiding the second-order density matrix	19
2.3.3 Integral estimates for AO-MP2 NMR shieldings	20
2.4 Spin component-scaled MP2 NMR shifts	22
3 Conclusion and outlook	23
4 Bibliography	25
5 Publications	29
5.1 Paper I: "A linear- and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Møller-Plesset perturbation theory", M. Maurer and C. Ochsenfeld, <i>J. Chem. Phys.</i> , 138 , 174104 (2013)	29

5.2	Paper II: "Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 levels", D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, C. Ochsenfeld, <i>J. Chem. Theory Comput.</i> , 10 , 572 (2014)	47
5.3	Paper III: "Spin component-scaled second-order Møller-Plesset perturbation theory for calculating NMR shieldings", M. Maurer and C. Ochsenfeld, <i>J. Chem. Theory Comput.</i> , 11 , 37 (2015)	57
5.4	Manuscript IV: "QQR-type integral estimates for calculating second-order Møller-Plesset perturbation theory-based NMR shieldings of selected nuclei with a sublinear-scaling computational effort", M. Maurer and C. Ochsenfeld, (in preparation)	97
5.5	Manuscript V: "QQR-type integral estimates for calculating linear-scaling energy gradients in atomic orbital-based second-order Møller-Plesset perturbation theory", M. Maurer and C. Ochsenfeld, (in preparation)	119
	Curriculum vitae	147

List of publications

This work is a cumulative dissertation, comprising five contributions in publication format. Three of these contributions are published in peer-reviewed journals (paper **I**, **II** and **III**). The last two contributions are currently in preparation and are therefore labelled as manuscripts **IV** and **V**:

Paper **I**: **Marina Maurer** and Christian Ochsenfeld, "A linear- and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Møller-Plesset perturbation theory", *J. Chem. Phys.*, **138**, 174104 (2013).

Paper **II**: Denis Flaig, **Marina Maurer**, Matti Hanni, Katharina Braunger, Leonhard Kick, Matthias Thubauville, and Christian Ochsenfeld, "Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 levels", *J. Chem. Theory Comput.*, **10**, 572 (2014).

Contribution by M. Maurer: *All derivations, the implementation, all calculations and most of the writing of the parts of this paper which deal with the GIAO-SCS- and GIAO-SOS-MP2 methods.*

Paper **III**: **Marina Maurer** and Christian Ochsenfeld, "Spin component-scaled second-order Møller-Plesset perturbation theory for calculating NMR shieldings", *J. Chem. Theory Comput.*, **11**, 37 (2015).

Manuscript **IV**: **Marina Maurer** and Christian Ochsenfeld, "QQR-type integral estimates for calculating second-order Møller-Plesset perturbation theory-based NMR shieldings of selected nuclei with a sublinear-scaling computational effort", *in preparation*.

Manuscript **V**: **Marina Maurer** and Christian Ochsenfeld, "QQR-type integral estimates for calculating linear-scaling energy gradients in atomic orbital-based second-order Møller-Plesset perturbation theory", *in preparation*.

Chapter 1

Introduction

Quantum chemical calculations of energies and properties of molecular systems are nowadays widely used in chemistry to support experimentally obtained data by providing useful complementary information. Since interest has increased over the years in studying biochemical molecules, which consist of several thousand atoms, calculations of energies and properties are increasingly important for such large systems. Unfortunately, the computational effort of conventional quantum chemical methods grows rapidly with increasing size of the system. Already the simplest quantum chemical method, the Hartree-Fock (HF) method [1, 2], scales formally with the fourth power of the molecule size. Therefore, much effort has been made to reduce the computational effort of the conventional quantum chemical methods. The computational cost of HF, for example, can be reduced to linear for large systems by applying the continuous fast multipole method (CFMM) [3, 4] and Link screening schemes [5, 6]. However, HF theory describes the electron-electron interactions only via the approximation of each electron interacting with the mean-field of all other electrons. The missing electron correlation effects can be included by post-HF methods, such as coupled-cluster (CC) [2], Møller-Plesset perturbation theory (MP) [1], or density-functional theory (DFT) [7]. The coupled-cluster method with single and double excitations and perturbative triples (CCSD(T)) provides very accurate results and is therefore often referred to as the "gold standard" in the field of quantum chemical methods. At the same time, the computational cost of CCSD(T) is formally increasing with the seventh power of the molecule size, which limits the calculations to very small molecules.

As an alternative approach, second-order Møller-Plesset perturbation theory (MP2) often provides accurate results with lower computational cost. Here, the missing electron correlation of HF theory is described by perturbation theory, where the sum of the zeroth- and first-order energy is again the HF energy. The first correction to the HF energy is therefore obtained at second order. Although the computational cost of the MP2 method increases formally with the fifth power of the molecule size, it can be reduced for larger molecules. Various approaches

have been introduced in the past to improve the efficiency of the MP2 method by combining it with the resolution-of-the-identity (RI) approximation [8–10] or by using local correlation approximations [11–13]. In this thesis, the focus is on the Laplace MP2 method, originally introduced by Almlöf and Häser [14–16], which allows to formulate the rate-determining steps in the local atomic-orbital (AO) basis that provides the foundation to reduce the computational effort. To exploit the local behavior of this method, integral estimates are used to preselect and only compute the significant contributions of the two-electron integrals, hence reducing the computational cost strongly. To this extent, S. A. Maurer et al. [17,18] introduced the so-called QQR-type estimates. These are based on the Schwarz estimates [19], but also include the dependence of the integral value on the distance between the charge distributions in bra and ket. By using these estimates, MP2 energies can be calculated with a linear-scaling behavior, which opens up the pathway for calculations of molecules with more than 1000 atoms and 10 000 basis functions on a single core computer [18].

For large systems, magnetic properties such as nuclear magnetic resonance (NMR) shieldings are also of special interest. These are obtained by differentiating the electronic energy with respect to the nuclear magnetic spin moment and the magnetic field. For HF- and DFT-based NMR shieldings, linear- and sublinear-scaling methods were already introduced with which systems with more than 1000 atoms can be studied on simple workstation computers [20,21]. While the results for MP2-based NMR shieldings are more reliable than those at the HF or DFT level, the formally high computational effort has to be reduced to allow for calculations of large systems. The efficiency of MO-MP2 NMR shieldings can be improved, e.g., with a local-correlation approximation as introduced by Gauss and Werner [22]. Recently, Loibl and Schütz [23] extended this idea by using a density fitting scheme. In the present work, the MP2 NMR shieldings are based on the AO-MP2 method. In paper I, the first MP2 NMR shielding method with a linear-scaling behavior for computing all shieldings and a sublinear-scaling behavior for a specific nucleus is introduced. The latter is most important, since in large systems, typically, not all the nuclei are of interest. For example, the NMR shieldings of the solvent molecules around a molecular system are not of interest and can therefore be avoided by using our method. The huge advantage of sublinear-scaling methods is that the computational cost no longer increases with the molecular size once a certain size has been reached, reducing the computational cost enormously. The sublinear-scaling behavior for a single nucleus can be achieved by exploiting the local behavior of the nuclear magnetic spin perturbation and avoiding the global magnetic field perturbation in the rate-determining steps. With a pilot implementation of the AO-MP2 NMR shieldings, the sublinear-scaling behavior of this method on linear alkanes as model systems is confirmed and the validity of the equations by comparing the results with MO-MP2 shieldings is shown. To ex-

exploit the sublinear-scaling behavior of the rate-determining steps, the QQR-type integral estimates were extended to NMR shieldings, as shown in manuscript **IV**. The preliminary results show the sublinear-scaling behavior for sufficiently large molecules and that differences between the AO-MP2 and MO-MP2 NMR shieldings can be fully controlled by a screening threshold. Furthermore, an AO-based formulation of the Z-vector method of Handy and Schaefer [24, 25] is developed to avoid the expensive calculation of the second derivative of the density matrix.

In most cases, the MP2 NMR shieldings are already quite accurate in comparison to other methods (see paper **II**), but their accuracy can be improved even further. As shown by Grimme [26] and others [27–29], the accuracy of energies can be improved by scaling the opposite and same spin parts of the MP2 correlation contribution differently. The so-called spin component scaled (SCS) MP2 method is quite beneficial for reaction energies, barrier heights, geometries, and harmonic vibrational frequencies [26]. Furthermore, Head-Gordon and coworkers [28] introduced the scaled opposite spin (SOS) MP2 method, which entirely neglects the same spin part and only scales the opposite spin part by another factor. The accuracy improvement of the SOS-MP2 method is similar to the SCS-MP2 method, but the great advantage is the improvement in the efficiency of this method. By combining the MO-MP2 method with the resolution-of-the-identity (RI) approximation and only determining the opposite spin part of the MP2 energy, the formal scaling behavior reduces from $\mathcal{O}(M^5)$ to $\mathcal{O}(M^4)$ [28]. Similar to these approaches, scaling factors are optimized in this work to improve the accuracy of MP2 NMR shieldings. Therefore, the carbon NMR shieldings of the benchmark set introduced in paper **II** is exploited, and new benchmark sets for other nuclei such as phosphorus, nitrogen, oxygen, and fluorine are presented. For all nuclei, the standard deviation is calculated with respect to CCSD(T)/cc-pVQZ results. As shown in papers **II** and **III**, our so-called GIAO-SCS-MP2 and GIAO-SOS-MP2 methods reduce the standard deviation for almost all presented basis sets in comparison to nonscaled MP2.

Besides NMR shieldings, also energy gradients for geometry optimization of large systems are important. In our group, Schweizer et al. [30] developed a theory for gradients within an AO-MP2 formulation, where all rate-determining steps can be calculated with a linear-scaling behavior. Schweizer et al. proved the validity of the equations with a pilot implementation of the AO-MP2 gradients. To exploit the linear-scaling nature of this theory, the QQR-type estimates were extended in the present work for energy gradients as summarized in manuscript **V**. The preliminary results show the desired asymptotic linear-scaling behavior for the rate-determining steps. Furthermore, the differences between the AO-MP2 gradients and the MO-MP2 gradients are determined, which are systematically dependent on the size of the screening thresholds.

Chapter 2

Theory

The most fundamental wave-function-based quantum-chemical method to describe a many-electron system is the Hartree-Fock (HF) theory. This method describes the wave-function of a system with a Slater determinant over one-electron functions, which leads to a mean field description of the electron-electron interactions [1]. With this approximation, one can obtain about 98 % of the total electronic energy. However, the missing 2 % is often very important for chemical questions. This so called electron correlation energy can be described with post-HF methods like MP2.

2.1 Møller-Plesset perturbation theory

2.1.1 Molecular orbital-based MP2 energy expression

The main idea behind the Møller-Plesset ansatz is to express the missing electron correlation of the HF method as a perturbation [1]. The perturbation operator is defined as the difference of the exact Hamiltonian and the Hamiltonian of the unperturbed system, which is described as a sum of Fock-operators. As a result, the energy of the unperturbed system is the sum of the orbital energies and adding the first-order correction, the HF energy results. Consequently, the first correction to the HF energy follows in the second order. The molecular-orbital (MO-) based second-order energy expression (MP2) for closed-shell molecules is:

$$E_{MP2} = - \sum_{ij}^{N_{occ}} \sum_{ab}^{N_{virt}} \frac{(ia|jb) [2(ia|jb) - (ib|ja)]}{-\varepsilon_i - \varepsilon_j + \varepsilon_a + \varepsilon_b} . \quad (2.1)$$

The indices i and j are describing occupied orbitals and the indices a and b virtual orbitals. The denominator consists of orbital energies ε .

2.1.2 Atomic orbital-based MP2 energy expression

The rate-determining step of the MO-MP2 method is the transformation of the two-electron integrals from the atomic orbital (AO) basis to the MO basis with the coefficient matrices:

$$(ia|jb) = \sum_{\mu}^N C_{\mu i} \left(\sum_{\nu}^N C_{\nu a} \left(\sum_{\lambda}^N C_{\lambda j} \left(\sum_{\sigma}^N C_{\sigma b} (\mu\nu|\lambda\sigma) \right) \right) \right) . \quad (2.2)$$

Each sum in 2.2 can be evaluated individually and has a computational cost of $\mathcal{O}(N^5)$, with N being the number of basis functions. Since the coefficient matrices \mathbf{C} are not local, the scaling behavior can not be reduced any further. Therefore, the energy expression must be formulated entirely in the AO basis for which the energy denominator has to be avoided. Almlöf and Häser [14–16] introduced the Laplace transformation to avoid the energy denominator:

$$\frac{1}{-\varepsilon_i - \varepsilon_j + \varepsilon_a + \varepsilon_b} = \int_0^{\infty} e^{-(\varepsilon_i - \varepsilon_j + \varepsilon_a + \varepsilon_b)t} dt \approx \sum_{\alpha=1}^{\tau} w^{(\alpha)} e^{\varepsilon_i t^{(\alpha)}} e^{\varepsilon_j t^{(\alpha)}} e^{-\varepsilon_a t^{(\alpha)}} e^{-\varepsilon_b t^{(\alpha)}} . \quad (2.3)$$

The Laplace transformation in Eq. 2.3 and the transformation from the AO to the MO basis in Eq. 2.2 are inserted in Eq. 2.1. After that, the occupied and unoccupied pseudo-density matrices can be defined

$$\begin{aligned} \underline{P}_{\mu'\mu} &= \sum_i C_{\mu' i} e^{\varepsilon_i t_{\alpha}} C_{\mu i} \\ \overline{P}_{\nu'\nu} &= \sum_a C_{\nu' a} e^{-\varepsilon_a t_{\alpha}} C_{\nu a} . \end{aligned} \quad (2.4)$$

As a result, the sum over the MO-indices i, j, a and b in Eq. 2.1 can be formed before the expensive two-electron integral transformation, which are now performed with these pseudo-densities:

$$(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma}) = \sum_{\mu'} \underline{P}_{\mu'\mu} \left(\sum_{\nu'} \overline{P}_{\nu'\nu} \left(\sum_{\lambda'} \underline{P}_{\lambda'\lambda} \left(\sum_{\sigma'} \overline{P}_{\sigma'\sigma} (\mu'\nu'|\lambda'\sigma') \right) \right) \right) , \quad (2.5)$$

The steps in Eq. 2.5 scale formally also with $\mathcal{O}(N^5)$, but in contrast to the coefficient matrices the pseudo-densities have an asymptotic linear-scaling number of significant elements, leading to a linear-scaling computational cost for larger molecules.

The AO-based MP2 energy is calculated by:

$$E_{\text{AO-MP2}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} E_{JK}^{(\alpha)} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\lambda\bar{\sigma}) [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] , \quad (2.6)$$

To reduce the error of the AO-MP2 energies below 1 kJ/mol for molecules with a significant HOMO-LUMO gap, one needs typically 5-6 Laplace points [18]. To reduce the scaling behavior of AO-MP2 calculations to linear, integral estimates are used, which are explained in the following section.

2.1.3 Integral estimates

Not all elements of the two-electron integrals are significant. The main idea of integral estimates is to preselect these significant contributions before the calculation of the two-electron integrals. By only calculating the significant two-electron integral the computational cost can be reduced, while the accuracy of the results can be controlled by a threshold. This can be done with the classical Schwarz estimates [19]:

$$|(\mu\nu|\lambda\sigma)| \leq \underbrace{|(\mu\nu|\mu\nu)|^{\frac{1}{2}}}_{Q_{\mu\nu}} \underbrace{|(\lambda\sigma|\lambda\sigma)|^{\frac{1}{2}}}_{Q_{\lambda\sigma}} . \quad (2.7)$$

These estimates exploit the following behavior of the two electron integrals: The number of two-electron integrals is formally N^4 (with N being the number of basis functions). Usually, Gaussian functions are used to describe the two-electron integrals, which have an exponential decay. In this case, there is only a constant number of basis functions ν around a basis functions μ , which results in a significant contribution. Therefore, by using the Schwarz estimates the scaling behavior of the number of required integrals for larger molecules can be reduced to $\mathcal{O}(N^2)$.

For AO-MP2 calculations, transformed integrals with pseudo-densities are calculated, as shown in Eq. 2.5. To estimate the contributions for these integrals, Häser [16] introduced the pseudo-Schwarz matrices, which are obtained by transforming the Schwarz estimates with pseudo-densities:

$$\begin{aligned} X_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\lambda\nu|\sigma\nu) \underline{P}_{\lambda\mu} \underline{P}_{\sigma\mu} \right)^{\frac{1}{2}} = (\underline{\mu}\nu|\underline{\mu}\nu)^{\frac{1}{2}} \\ Y_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\mu\lambda|\mu\sigma) \bar{P}_{\lambda\nu} \bar{P}_{\sigma\nu} \right)^{\frac{1}{2}} = (\mu\bar{\nu}|\mu\bar{\nu})^{\frac{1}{2}} \\ Z_{\mu\nu} &= \min_{\lambda} \left(\sum_{\lambda} X_{\mu\lambda} |\bar{P}_{\lambda\nu}| , \sum_{\lambda} |P_{\mu\lambda}| Y_{\lambda\nu} \right) \geq (\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})^{\frac{1}{2}} . \end{aligned} \quad (2.8)$$

The Schwarz estimates are exploiting the exponential coupling between the basis functions μ and ν and also between λ and σ of the two-electron integral in Eq. 2.7. However, the Schwarz estimates neglect the dependence of the integral value on the distance between the charge contributions in bra and ket. This distance is included in the recently introduced QQR-type estimates [17, 18]. By using these estimates, the number of preselected two-electron integrals can be reduced even further than with the Schwarz-estimates such that the scaling behavior is reduced from quadratic to linear for larger molecules.

For the QQR-type estimates, the decay behavior of the two-electron integrals between the contributions in bra and ket has to be determined. Therefore, the multipole expansion is used. The first terms of the multipole expansion are

$$\begin{aligned}
 (\Omega_A|\Omega_B) = & \frac{q_{00}^A q_{00}^B}{R_{AB}} + \frac{q_{00}^A \left(\sum_{j=-1}^1 T'_{00,1j} q_{1j}^B \right)}{R_{AB}^2} \\
 & + \frac{\left(\sum_{i=-1}^1 q_{1i}^A T'_{1i,00} \right) q_{00}^B}{R_{AB}^2} + \mathcal{O}(R_{AB}^{-3}) ,
 \end{aligned} \tag{2.9}$$

where the basis functions in bra and ket are described by a charge distribution Ω . The first term shows the monopole-monopole interactions, the second term the monopole-dipole interactions and so on.

If we are expanding a transformed integral with pseudo-densities, the monopoles in the multipole expansion in Eq. 2.9 are zero, because of the orthogonality of the occupied and virtual subspace:

$$q_{00}^{\underline{\mu}\bar{\nu}} = S_{\underline{\mu}\bar{\nu}} = \sum_{\mu'\nu'} \underline{P}_{\mu'\mu} S_{\mu\nu} \bar{P}_{\nu\nu'} = 0 \tag{2.10}$$

The AO-MP2 energy in Eq. 2.6 can also be calculated as the contraction of two half-transformed integrals instead of a fully-transformed integral with an untransformed integral [18]. Using the multipole expansion, the QQR-type estimate for a half-transformed integral is

$$(\underline{\mu}\bar{\nu}|\lambda\sigma) \approx \frac{Z_{\underline{\mu}\nu} Q_{\lambda\sigma}}{R'_{\underline{\mu}\bar{\nu},\lambda\sigma}{}^2} \tag{2.11}$$

Here, the Schwarz matrices \mathbf{Z} and \mathbf{Q} are used. The asymptotic decay of this half-transformed integral is $1/(R')^2$, because the first and second term in the multipole expansion in Eq. 2.9 vanishes since the monopoles $q_{00}^{\underline{\mu}\bar{\nu}}$ are zero. As a result, the third term with a decay of $1/(R')^2$ determines the asymptotic decay behavior.

The distance R' of bra and ket is calculated by the difference of the distance R of the two centers of bra and ket and the extents (ext) of the respective charge distributions:

$$R'_{\underline{\mu}\bar{\nu},\lambda\sigma} = R - \text{ext}_{\underline{\mu}\bar{\nu}} - \text{ext}_{\lambda\sigma} \quad (2.12)$$

The multipole expansion in Eq. 2.9 can only be applied if the charge distributions in bra Ω_A and ket Ω_B are well-separated. Therefore, the QQR-type estimates are only applied, if R' is greater than one. Otherwise, the conventional Schwarz estimates \mathbf{Z} and \mathbf{Q} are used.

By using the QQR-type estimates within AO-MP2, MP2 energies can be calculated with a linear-scaling computational cost for molecules with a significant HOMO-LUMO gap [18]. As a result, calculations on non-metallic systems with more than 1000 atoms and 10 000 basis functions are accessible on a single core computer [18].

2.2 AO-based MP2 energy gradients

The first derivative of the energy with respect to the nuclear coordinates are the energy gradients and can be used to optimize the structure of a molecule. The theory and a preliminary implementation of the following AO-MP2 gradients was presented by Schweizer et al. in 2008 [30]. Due to the entirely AO-based formulation of the rate-determining steps, the method opens the way to calculate MP2 gradients of larger molecules with a linear-scaling computational cost. To exploit the asymptotic linear-scaling behavior, we developed an extension to the QQR-type estimates, originally introduced for HF and AO-MP2 energies [17, 18]. In the following, a short summary of the theory of the AO-MP2 gradients [30] and then the QQR-type estimates for this method are presented.

2.2.1 The first derivative with respect to the nuclear coordinates ξ

As shown by Schweizer et al. [30], the first derivative of the AO-MP2 energy expression in Eq. 2.6 with respect to the nuclear coordinates ξ is given by:

$$\begin{aligned} E_{\text{AO-MP2}}^{\xi} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2 \sum_{\underline{\mu}\bar{\nu}\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma}) (\underline{\mu}\bar{\nu}||\lambda\sigma)^{\xi} \\ & + 2 \sum_{\underline{\mu}\bar{\nu}\lambda\sigma} \sum_{\mu'} P_{\mu'\mu}^{\xi} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma}) (\underline{\mu}\bar{\nu}||\lambda\sigma) \\ & + 2 \sum_{\underline{\mu}\bar{\nu}\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\xi} (\underline{\mu}\bar{\nu}'|\underline{\lambda}\bar{\sigma}) (\underline{\mu}\bar{\nu}||\lambda\sigma) \} \end{aligned} \quad (2.13)$$

After defining the following matrices

$$\begin{aligned}\mathcal{I}^\xi &= \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma)^\xi \\ \bar{R}_{\mu'\mu} &= \sum_{\nu\lambda\sigma} (\mu'\bar{\nu}|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma) \\ \underline{R}_{\nu'\nu} &= \sum_{\mu\lambda\sigma} (\underline{\mu}\nu'|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma) \ ,\end{aligned}\tag{2.14}$$

the first derivative of the energy can be written as:

$$E_{\text{AO-MP2}}^\xi = - \sum_{\alpha=1}^{\tau} w_\alpha \left\{ 2 \left[\mathcal{I}^\xi + \sum_{\mu'\mu} \bar{R}_{\mu'\mu} \underline{P}_{\mu'\mu}^\xi + \sum_{\nu'\nu} \underline{R}_{\nu'\nu} \bar{P}_{\nu'\nu}^\xi \right] \right\} . \tag{2.15}$$

The calculation of the \mathcal{I}^ξ term, containing the derivatives of the two-electron integrals, is not rate-determining. In contrast, the determination of the explicit derivative of the pseudo-densities \underline{P}^ξ and \bar{P}^ξ is quite expensive. However, the explicit calculation of them can be avoided by using the Z-vector method of Handy and Schaefer [24, 25], which was reformulated in the AO basis by Schweizer et al. [30]. To be able to apply the Z-vector method, the AO-MP2 gradient equations in 2.15 has to be reformulated as shown in the next section.

2.2.2 Avoiding the computation of the perturbed pseudo-densities

First, the derivatives of the pseudo-densities are shown to explain the expensive steps in calculating them. In Eq. 2.4, the pseudo-densities are calculated with the coefficient matrices \mathbf{C} , but they can also be obtained with a linear-scaling computational cost using alternatively the conventional occupied \mathbf{P}_{occ} and unoccupied one-particle density matrices \mathbf{P}_{virt} and the Fock matrix \mathbf{F} [30–32]:

$$\underline{P}_{\mu'\mu} = \sum_i C_{\mu'i} e^{\varepsilon_i t_\alpha} C_{\mu i} = (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}})_{\mu'\mu} \tag{2.16a}$$

$$\bar{P}_{\nu'\nu} = \sum_a C_{\nu'a} e^{-\varepsilon_a t_\alpha} C_{\nu a} = (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}})_{\nu'\nu} . \tag{2.16b}$$

The corresponding derivatives with respect to the nuclear coordinates ξ of these pseudo-densities are:

$$\begin{aligned}\underline{P}_{\mu'\mu}^\xi &= (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^\xi \mathbf{P}_{\text{occ}} + e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}}^\xi \\ \bar{P}_{\nu'\nu}^\xi &= (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^\xi \mathbf{P}_{\text{virt}} + e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}}^\xi \ ,\end{aligned}\tag{2.17}$$

where $e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}}$, $e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}}$ and the analogous derivatives are matrix exponentials and can be calculated by a Taylor series [30]. Inserting these derivatives of the pseudo-densities in Eq. 2.15 results in:

$$E_{\text{AO-MP2}}^\xi = - \sum_{\alpha=1}^{\tau} w_\alpha \left\{ 2 \mathcal{I}^\xi + 2 \text{Tr} \left[\underbrace{\overline{\mathbf{R}} \left((e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^\xi \mathbf{P}_{\text{occ}} \right)}_{\mathcal{A}_1} + \underbrace{e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}}^\xi}_{\mathcal{B}_1} \right] \right. \\ \left. + 2 \text{Tr} \left[\underbrace{\overline{\mathbf{R}} \left((e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^\xi \mathbf{P}_{\text{virt}} \right)}_{\mathcal{A}_2} + \underbrace{e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}}^\xi}_{\mathcal{B}_2} \right] \right\} . \quad (2.18)$$

The expensive step of determining the derivative of the pseudo-densities, is the calculation of occupied $\mathbf{P}_{\text{occ}}^\xi$ and virtual $\mathbf{P}_{\text{virt}}^\xi$ densities. The derivative of the virtual density-matrix $\mathbf{P}_{\text{virt}}^\xi$ can be substituted by an expression including $\mathbf{P}_{\text{occ}}^\xi$ by exploiting the relation, which is valid for a non-orthogonal basis:

$$\mathbf{P}_{\text{occ}} \mathbf{S} + \mathbf{P}_{\text{virt}} \mathbf{S} = \mathbf{1} . \quad (2.19)$$

After differentiating this relation and multiplying from the right with \mathbf{S}^{-1} , one obtains

$$\mathbf{P}_{\text{virt}}^\xi = -\mathbf{P}_{\text{occ}}^\xi - \underbrace{(\mathbf{P}_{\text{occ}} + \mathbf{P}_{\text{virt}})}_{\mathbf{S}^{-1}} \mathbf{S}^\xi \mathbf{S}^{-1} . \quad (2.20)$$

To be able to avoid the explicit calculation of $\mathbf{P}_{\text{occ}}^\xi$ by applying the Z-vector-method of Handy and Schaefer [24, 25], the AO-MP2 gradient equations must have a special form:

$$E_{\text{AO-MP2}}^\xi = \text{Tr} [\mathcal{P} \mathbf{P}_{\text{occ}}^\xi] + \text{Tr} [\mathcal{X}] , \quad (2.21)$$

All the terms depending on $\mathbf{P}_{\text{occ}}^\xi$ are in the first term and the rest in the second. Furthermore, the perturbed density matrix has to be on the far right hand side within the trace.

While the perturbed density matrices in the terms \mathcal{B}_1 and \mathcal{B}_2 in Eq. (2.18) are already on the far right hand side, the terms \mathcal{A}_1 and \mathcal{A}_2 have to be rearranged. For this rearrangement, the expansions of the exponentials are inserted and cyclic permutations within the trace are applied. For example for the \mathcal{A}_1 term, it results:

$$\text{Tr} [\overline{\mathbf{R}} (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^\xi \mathbf{P}_{\text{occ}}] = \text{Tr} [(\overline{\mathbf{Y}}_1 + \mathbf{G}[\overline{\mathbf{Y}}_2]) \mathbf{P}_{\text{occ}}^\xi] + \text{Tr} [\overline{\mathbf{Y}}_2 \mathbf{F}^{(\xi)}] . \quad (2.22)$$

Here $\bar{\mathbf{Y}}_1 = \sum_{k=1}^m \bar{\mathbf{Y}}_1^{(k)}$ and $\bar{\mathbf{Y}}_2 = \sum_{k=1}^m \bar{\mathbf{Y}}_2^{(k)}$ are the recursion formulae, where each term is calculated using

$$\bar{\mathbf{Y}}_1^{(k)} = \frac{1}{k} \left[(t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}} \bar{\mathbf{R}}) (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(k-1)} + (t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}}) \bar{\mathbf{Y}}_1^{(k-1)} \right] \quad (2.23a)$$

$$\bar{\mathbf{Y}}_2^{(k)} = \frac{1}{k} \left[(t_\alpha \mathbf{P}_{\text{occ}} \bar{\mathbf{R}} \mathbf{P}_{\text{occ}}) (t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}})^{(k-1)} + (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}) \bar{\mathbf{Y}}_2^{(k-1)} \right], \quad (2.23b)$$

The recursions start with $\bar{\mathbf{Y}}_1^{(0)} = 0$ and $\bar{\mathbf{Y}}_2^{(0)} = 0$, respectively. In Eq. 2.22, the derivative of the Fock-matrix is split into two parts. One part is depending on the perturbed density matrix and the other parts are included in $\mathbf{F}^{(\xi)}$:

$$\mathbf{F}^\xi = \underbrace{\mathbf{h}^\xi + \mathbf{G}^\xi [\mathbf{P}_{\text{occ}}]}_{\mathbf{F}^{(\xi)}} + \mathbf{G} [\mathbf{P}_{\text{occ}}^\xi]. \quad (2.24)$$

Here, \mathbf{G} is used as an abbreviation for the two-electron integrals. For example $\mathbf{G}[\mathbf{P}_{\text{occ}}^\xi]$ represents the contraction of the two-electron integrals with the perturbed density matrix. For the other term \mathcal{A}_2 , the result is similar to the \mathcal{A}_1 term (for further details see Ref. [30]).

Collecting all the terms, the first derivative with respect to the nuclear coordinates can be formulated as:

$$\begin{aligned} E_{\text{AO-MP2}}^\xi &= - \sum_{\alpha=1}^{\tau} w_\alpha \{ 2 \mathcal{I}^\xi \\ &\quad + 2 \text{Tr} \left[\left(\bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G} [\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] + \bar{\mathbf{R}} e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} - \underline{\mathbf{R}} e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \right) \mathbf{P}_{\text{occ}}^\xi \right] \\ &\quad + 2 \text{Tr} \left[\left(\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2 \right) \mathbf{F}^{(\xi)} \right] \\ &\quad + 2 \text{Tr} \left[- \left(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}} e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \right) \mathbf{S}^{-1} \mathbf{S}^\xi \mathbf{S}^{-1} \right] \} \\ &= - \sum_{\alpha=1}^{\tau} w_\alpha \{ 2 \mathcal{I}^\xi + 2 \text{Tr} [\mathcal{P} \mathbf{P}_{\text{occ}}^\xi] + 2 \text{Tr} [\mathcal{F} \mathbf{F}^{(\xi)}] + 2 \text{Tr} [\mathcal{S} \mathbf{S}^{-1} \mathbf{S}^\xi \mathbf{S}^{-1}] \} \end{aligned} \quad (2.25)$$

with

$$\mathcal{P} = \bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G} [\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] + \bar{\mathbf{R}} e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} - \underline{\mathbf{R}} e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \quad (2.26a)$$

$$\mathcal{F} = \bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2 \quad (2.26b)$$

$$\mathcal{S} = - \left(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}} e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \right). \quad (2.26c)$$

Here, all the terms depending on the perturbed density matrix are included in $\text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^\xi]$ as required for the Z-vector method in Eq. 2.21.

2.2.3 AO-based Z-vector method for avoiding the first-order density matrix

In general, perturbed density matrices are calculated with coupled-perturbed self-consistent field (CPSCF) methods. The equations of these methods for the here required perturbed density matrix $\mathbf{P}_{\text{occ}}^\xi$ would be dependent on $3 \cdot N_{\text{atoms}}$ perturbations. The advantage of using the Z-vector method of Handy and Schaefer [24, 25] is that the equations are not dependent on perturbations and therefore only one CPSCF related equation has to be solved. However, the original Z-vector method is formulated in the MO basis and no linear-scaling implementation is possible. For this reason, Schweizer et al. [30] reformulated the Z-vector method into the AO basis.

We first start with the abbreviated density-matrix-based CPSCF (D-CPSCF) equations (see also appendix C of paper I):

$$\text{Tr}[\underline{\underline{\mathbf{A}}} \mathbf{P}_{\text{occ}}^\xi] = \text{Tr}[\mathbf{b}^\xi] . \quad (2.27)$$

$\underline{\underline{\mathbf{A}}}$ is a symmetric positive-definite Hessian, which can be inverted $\underline{\underline{\mathbf{A}}}^{-1}$ and multiplied from the left:

$$\text{Tr}[\mathbf{P}_{\text{occ}}^\xi] = \text{Tr}[\underline{\underline{\mathbf{A}}}^{-1} \mathbf{b}^\xi] . \quad (2.28)$$

To get an expression for $\text{Tr}[\mathcal{P} \mathbf{P}^\xi]$, the final step is to multiply from the left with \mathcal{P} and the Z-vector can be defined as:

$$\begin{aligned} \text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^\xi] &= \text{Tr}[\mathcal{P} \underline{\underline{\mathbf{A}}}^{-1} \mathbf{b}^\xi] . \\ &= \underbrace{\mathbf{Z}^T} \end{aligned} \quad (2.29)$$

As a result, the term $\text{Tr}[\mathcal{P} \mathbf{P}^\xi]$ in Eq. 2.25 can be substituted with $\text{Tr}[\mathbf{Z}^T \mathbf{b}^\xi]$. The resulting Z-vector \mathbf{Z} can be determined with a D-CPSCF related equation:

$$\underline{\underline{\mathbf{A}}} \mathbf{Z} = \mathcal{P} . \quad (2.30)$$

This determination of the Z-vector is not dependent on any perturbations in comparison to the required calculations for $\mathbf{P}_{\text{occ}}^\xi$ in Eq. 2.27 and therefore the Z-vector can be obtained in a much more efficient way.

2.2.4 Integral estimates for AO-MP2 gradients

The most expensive steps in calculating AO-MP2 gradients are the determination of the integral products in Eqs. 2.14:

$$(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma)^\xi, (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) \text{ and } (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma). \quad (2.31)$$

To calculate only the significant integral products of these terms, we developed an extension to the QQR-type estimates, which were introduced by S. A. Maurer et al. for HF and AO-MP2 energies [17, 18]. As discussed in section 2.1.3, for the QQR-type estimates one uses the Schwarz and pseudo-Schwarz estimates and the decay depending on the distance of the bra and ket contributions. To determine the actual decay behavior of the integral products, the multipole expansion is exploited. For the AO-MP2 energies, half-transformed integrals are estimated as shown in section 2.1.3. For the AO-MP2 gradients, we need estimates for fully-transformed and triply-transformed integrals, as shown in Eq. 2.31. For the QQR-type estimate for the first term, we obtain:

$$(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma)^\xi \approx \frac{Z_{\underline{\mu}\bar{\nu}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^3(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (2.32)$$

For the derivative of the two-electron integrals, the estimates of the non-perturbed two-electron integrals can be used [6, 33]. The matrices \mathbf{Q} and \mathbf{Z} are the previously introduced Schwarz and pseudo-Schwarz estimates as defined in Eq. 2.7 and 2.8, respectively. For the fully-transformed integrals using the multipole expansion of Eq. 2.9, we can determine an asymptotic decay behavior of $1/(R')^3$, since the monopoles $q_{00}^{\underline{\mu}\bar{\nu}}$ and $q_{00}^{\underline{\lambda}\bar{\sigma}}$ are both zero. Therefore, the first three terms of the multipole expansion are zero and the fourth term with a decay of $1/(R')^3$ determines the decay behavior. For an untransformed integral, none of the terms vanish and a decay behavior of $1/R'$ is observed.

The significant contributions of the other two integral products in Eq. 2.31, including triply-transformed integrals, can be estimated by:

$$\begin{aligned} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) &\approx \frac{Y_{\mu'\bar{\nu}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\mu'\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \\ (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) &\approx \frac{X_{\underline{\mu}\nu'}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})}. \end{aligned} \quad (2.33)$$

\mathbf{Y} , \mathbf{X} and \mathbf{Z} are again the pseudo-Schwarz matrices and \mathbf{Q} is the Schwarz matrix. For the triply-transformed integrals, a decay behavior of $1/(R')^2$ results, since the

monopoles $q_{00}^{\lambda\sigma}$ are zero. Therefore, the first term in the multipole expansion in Eq. 2.9 is zero and the second term determines the decay behavior.

These estimates for the AO-MP2 gradients are also discussed in manuscript **V**. The preliminary results for applying the QQR-type estimates to the AO-MP2 gradients show the asymptotic linear-scaling behavior for the number of integral products in Eq. 2.31. The deviations between the AO-MP2 gradients compared to the MO-MP2 gradients can thereby be controlled by a screening threshold. Furthermore, the total wall times of the AO-MP2 and MO-MP2 gradients methods are compared. The speedup of an AO-MP2 gradient calculation of a DNA double-strand with four base pairs (STO-3G basis set) is roughly a factor of 18 compared to a conventional MO-MP2 gradient calculation. While these are preliminary timings obtained using the STO-3G basis, it is shown in manuscript **V** that for linear alkanes the timing ratios for STO-3G and 6-31G* are similar, so that the ratio seems at first sight reasonable.

Furthermore, based on the success of the AO-MP2 gradients, we are currently also developing in our group a method, which combines this ansatz with the resolution-of-the-identity (RI) approximation and the Cholesky decomposition of the pseudo-densities (CDD) similar to RI-CDD-energies shown by S.A. Maurer et al. [34]. This approach shows for MP2 energies huge efficiency improvements for calculations with large basis sets and is therefore quite promising for the use in gradient calculations.

2.3 AO-based MP2 NMR shieldings

NMR shieldings can be obtained by calculating the second derivative of the energy with respect to the nuclear magnetic spin moment and the magnetic field. Since MP2-based NMR shieldings have proven to provide quite accurate results (see paper **II**), it is important to overcome the formally steep scaling of $\mathcal{O}(N^5)$ of the MP2 method to open the way for calculations of large molecules on the MP2 level. Therefore, we developed the first theory for determining NMR shieldings at the MP2 level with a linear computational cost or even sublinear, if single nuclei are investigated. As shown in paper **I**, we focus on the calculation of selected nuclei, because we aim to calculate larger molecules, where often just a few nuclei are of interest. The sublinear-scaling for one nucleus can be achieved by exploiting local quantities perturbed with respect to the nuclear magnetic spin moment. As shown in figure 2.1, the perturbed pseudo-density matrix $\underline{P}^{\mathbf{m}_j}$, for example, has only a sublinear-scaling number of significant elements in comparison to the linear-scaling perturbed density matrix $\mathbf{P}^{\mathbf{B}}$ with respect to the global magnetic field perturbation. Therefore, the theory has to be formulated exploiting this behavior, which is extensively discussed in paper **I** and briefly summarized in the next sections.

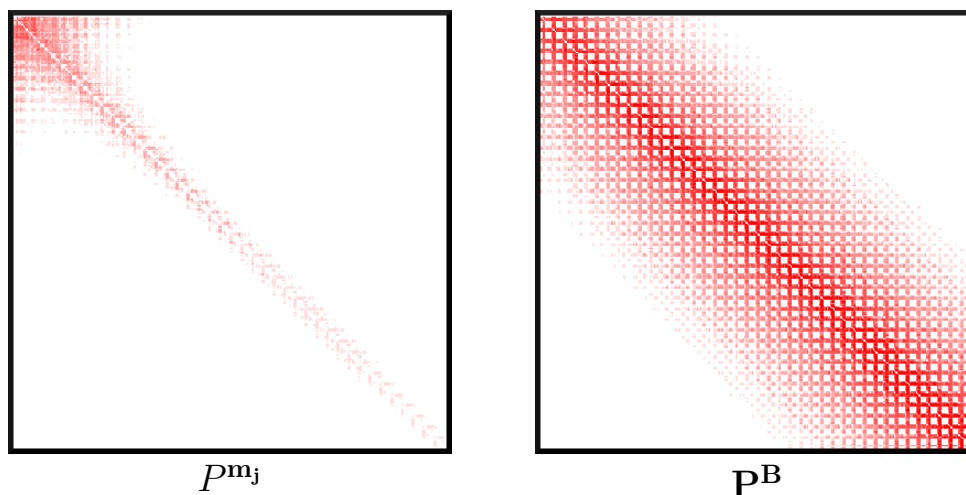


Figure 2.1: Sparsity pattern of the perturbed pseudo-density matrix with respect to the nuclear magnetic spin moment $\underline{P}^{\mathbf{m}_j}$ and the perturbed density matrix with respect to the magnetic field $\mathbf{P}^{\mathbf{B}}$ for a linear alkane with 40 carbon atoms (basis set: STO-3G). Depending on the size of the element of the matrix, the elements are signed by different color. Matrix elements larger than 10^{-5} are shown in different shades of red.

2.3.1 The AO-MP2-based magnetic shielding tensor

The second derivative of the energy with respect to the magnetic field \mathbf{B} and the nuclear magnetic spin moment \mathbf{m}_j results in the magnetic shielding tensor σ_j . To obtain the AO-MP2-based shielding tensor, the AO-MP2 energy expression in Eq. 2.6 is differentiated with respect to these perturbations. Each step is shown in paper I, where the AO-based shielding tensor results in:

$$\begin{aligned}
 \sigma_j^{MP2} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 4 \sum_{\mu\nu\lambda\sigma} \sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'\bar{\nu}|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}} \\
 & + 4 \sum_{\mu\nu\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\mu\nu'|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}} \\
 & + 2\text{Tr} \left[(\bar{\mathbf{Y}}_1^{\mathbf{m}_j} - \underline{\mathbf{Y}}_1^{\mathbf{m}_j} + \mathbf{G}[\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \underline{\mathbf{Y}}_2^{\mathbf{m}_j}] + \bar{\mathbf{R}}^{\mathbf{m}_j} e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} \right. \\
 & \quad \left. + \bar{\mathbf{R}} (e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}})^{\mathbf{m}_j} - \underline{\mathbf{R}}^{\mathbf{m}_j} e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} \right. \\
 & \quad \left. - \underline{\mathbf{R}} (e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j} \right) \mathbf{P}_{\text{occ}}^{\mathbf{B}} \Big] \\
 & + 2\text{Tr} \left[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j} \right] \\
 & + 2\text{Tr} \left[(\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \underline{\mathbf{Y}}_2^{\mathbf{m}_j}) \mathbf{F}^{(\mathbf{B})} + (\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2) \mathbf{F}^{(\mathbf{Bm}_j)} \right] \\
 & + 2\text{Tr} \left[-(\underline{\mathbf{Y}}_1^{\mathbf{m}_j} + \underline{\mathbf{R}}^{\mathbf{m}_j} e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} \right. \\
 & \quad \left. + \underline{\mathbf{R}} (e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j}) \mathbf{S}^{-1} \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1} \right] \} .
 \end{aligned} \tag{2.34}$$

Here, the second derivative of the density-matrix $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$ appears and the first derivative of the density matrix perturbed with respect to the magnetic field $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$. For a fully sublinear-scaling method, the explicit calculation of these matrices must be avoided. Therefore, we developed an AO-based reformulation of the Z-vector method of Handy and Schaefer [24, 25]. The equations for avoiding the explicit calculation of $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ is shown in paper I. For avoiding the second order derivative of the density matrix $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$, the AO-based Z-vector equations are presented in the next section.

2.3.2 AO-based Z-vector method for avoiding the second-order density matrix

The second-order density matrix of the term $\text{Tr} \left[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j} \right]$ in Eq. 2.34 is dependent on $9 \cdot N_{\text{atoms}}$ perturbations. For efficient calculations, we avoid the explicit calculation of $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$ by reformulating the Z-vector method in the AO basis. Therefore,

we start with the abbreviated D-CPSCF equations [20, 35] (for more information see also appendix C of paper **I**):

$$\text{Tr}[\underline{\underline{\mathbf{A}}}\mathbf{P}_{\text{occ}}^{\mathbf{B}}] = \text{Tr}[\mathbf{b}^{\mathbf{B}}] , \quad (2.35)$$

For the second derivative, the equation is differentiated with respect to the nuclear magnetic spin moment \mathbf{m}_j :

$$\text{Tr}[\underline{\underline{\mathbf{A}}}\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j} + \underline{\underline{\mathbf{A}}}^{\mathbf{m}_j}\mathbf{P}_{\text{occ}}^{\mathbf{B}}] = \text{Tr}[\mathbf{b}^{\mathbf{Bm}_j}] . \quad (2.36)$$

After multiplying this expression from the left with the inverse of the symmetric, positive-definite Hessian $\underline{\underline{\mathbf{A}}}$ and after that with the matrix \mathcal{P} , the Z-vector can be defined as:

$$\text{Tr}[\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}] = \underbrace{\text{Tr}[\mathcal{P}\underline{\underline{\mathbf{A}}}^{-1}(\mathbf{b}^{\mathbf{Bm}_j} - \underline{\underline{\mathbf{A}}}^{\mathbf{m}_j}\mathbf{P}_{\text{occ}}^{\mathbf{B}})]}_{=\mathbf{Z}^T} , \quad (2.37)$$

As a result, the term $\text{Tr}[\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}]$ can be substituted with $\text{Tr}[\mathbf{Z}^T(\mathbf{b}^{\mathbf{Bm}_j} - \underline{\underline{\mathbf{A}}}^{\mathbf{m}_j}\mathbf{P}_{\text{occ}}^{\mathbf{B}})]$. In contrast to the $9N_{\text{atoms}}$ perturbations of $\mathbf{P}^{\mathbf{Bm}_j}$, the Z-vector \mathbf{Z}^T is independent on perturbations and can therefore be efficiently calculated. To determine the Z-vector, we are using the efficient Laplace-based D-CPSCF equations (DL-CPSCF) [36], as shown in appendix D of paper **I**.

The AO-MP2-based shielding tensor was first implemented in a simple fashion to prove the theory and to investigate the asymptotic scaling behavior. In paper **I**, the results are presented, which show the sublinear-scaling behavior for linear alkanes as model systems and the accuracy in comparison to MO-based shieldings in dependence of the number of Laplace points.

As a next step, integral estimates were developed to exploit the sublinear-scaling behavior for larger molecules, which is shown in manuscript **IV** and briefly in the next section.

2.3.3 Integral estimates for AO-MP2 NMR shieldings

We present in manuscript **IV** the integral estimates for the AO-MP2 NMR shieldings, which are an extension to the QQR-type estimates introduced for energies [17, 18]. The different estimates for the integral products needed for the AO-MP2 NMR shieldings, are discussed in manuscript **IV**. For example, for the $\underline{\underline{\mathbf{R}}}_{\nu'\nu}^{\mathbf{m}_j}$ matrices in Eq. 2.34, the following integral products are estimated:

$$(\underline{\underline{\mu}}^{\mathbf{m}_j}\nu'|\underline{\underline{\lambda}}\bar{\sigma})(\mu\nu|\lambda\sigma), (\underline{\underline{\mu}}\nu'|\underline{\underline{\lambda}}^{\mathbf{m}_j}\bar{\sigma})(\mu\nu|\lambda\sigma) \text{ and } (\underline{\underline{\mu}}\nu'|\underline{\underline{\lambda}}\bar{\sigma}^{\mathbf{m}_j})(\mu\nu|\lambda\sigma).$$

As discussed in section 2.1.3, the Schwarz and pseudo-Schwarz estimates are exploited within the QQR-type estimates. Since we have in Eq. 2.3.3 transformed

integrals with perturbed pseudo-densities with respect to the nuclear magnetic spin moment, we transform the Schwarz matrices with these perturbed pseudo-densities:

$$\begin{aligned} X_{\underline{\mu}^{\mathbf{m}_j\nu}} &= \left(\sum_{\lambda\sigma} (\lambda\nu|\sigma\nu) P_{\lambda\mu}^{\mathbf{m}_j} P_{\sigma\mu}^{\mathbf{m}_j} \right)^{\frac{1}{2}} = (\underline{\mu}^{\mathbf{m}_j\nu}|\underline{\mu}^{\mathbf{m}_j\nu})^{\frac{1}{2}} \\ Z_{\underline{\mu}\bar{\nu}^{\mathbf{m}_j}} &= \left(\sum_{\lambda} X_{\mu\lambda} |\bar{P}_{\lambda\nu}^{\mathbf{m}_j}| \right) \\ Z_{\underline{\mu}^{\mathbf{m}_j}\bar{\nu}} &= \left(\sum_{\lambda} |P_{\mu\lambda}^{\mathbf{m}_j}| Y_{\lambda\nu} \right) . \end{aligned} \quad (2.38)$$

Furthermore, the QQR-type estimates include the decay behavior of the contributions in bra and ket, which is determined by analyzing the multipole expansion of the integral, as shown in section 2.1.3.

For the first integral product in Eq. 2.3.3, we obtain the following QQR-type estimate:

$$(\underline{\mu}^{\mathbf{m}_j\nu'}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) \approx \frac{X_{\underline{\mu}^{\mathbf{m}_j\nu'}} Z_{\underline{\lambda}\bar{\sigma}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}^{\mathbf{m}_j\nu'}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2 (R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} . \quad (2.39)$$

By exploiting the multipole expansion, a decay behavior of $1/(R')^2$ results for the triply-transformed integral, because the monopoles $q_{00}^{\lambda\bar{\sigma}}$ are zero.

For the triply-transformed integrals of the last two integral products in Eq. 2.3.3, the decay behavior is $1/R'$:

$$\begin{aligned} (\underline{\mu}\nu'|\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma})(\mu\nu|\lambda\sigma) &\approx \frac{X_{\underline{\mu}\nu'} Z_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \\ (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j})(\mu\nu|\lambda\sigma) &\approx \frac{X_{\underline{\mu}\nu'} Z_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} . \end{aligned} \quad (2.40)$$

The preliminary results for applying the QQR-type estimates for the AO-MP2 shieldings are presented in manuscript **IV**. The computational cost of the rate-determining steps of AO-MP2 NMR calculations of a specific nucleus can thereby be reduced to sublinear for large molecules. Furthermore, the deviations between the AO-MP2 shieldings and the MO-MP2 shieldings are determined for different screening thresholds, which are systematically dependent on the size of the thresholds.

2.4 Spin component-scaled MP2 NMR shifts

The AO-MP2 shieldings, presented in this work, opens the way to calculate larger molecules at the MP2 level. Furthermore, we developed a method to improve the general accuracy of MP2 NMR shieldings by scaling the same and opposite spin-components of MP2 NMR shieldings differently, similar to the scaling schemes introduced for MP2 energies by Grimme [26] and others [27–29], as shown in papers **II** and **III**. Therefore, we split the NMR shielding tensor into the HF contribution and the correlation part into the opposite spin (OS) and same spin (SS) terms:

$$\sigma_{\text{SCS-MP2}} = \sigma_{\text{HF}} + c_{\text{OS}} \cdot \sigma_{\text{OS}} + c_{\text{SS}} \cdot \sigma_{\text{SS}} . \quad (2.41)$$

To determine the scaling factors c_{OS} and c_{SS} for carbon NMR shifts, we employ the benchmark set introduced by Flaig et al. in paper **II**. This work presents the accuracy of carbon and hydrogen NMR shifts of various organic molecules calculated with CCSD(T), MP2 and different DFT functionals. Within this study, the results are compared to CCSD(T)/cc-pVQZ data, since they are the most reliable theoretical NMR data available for this benchmark set. Following this approach, additional structures were optimized to study also the accuracy of scaled MP2 shifts for phosphorus, nitrogen, oxygen, and fluorine nuclei (see paper **III**). For all different nuclei, we optimized scaling factors in a least-square procedure to the CCSD(T)/cc-pVQZ results for each single basis set (the explicit equations for the fitting procedure are shown in paper **III**). We called this approach GIAO-SCS-MP2 method. Furthermore, as shown for energies by Head-Gordon and coworkers [28] the same-spin part can also be entirely neglected and only the opposite spin part is scaled by a factor, which saves computational time. For this approach, we also optimized new scaling factors for MP2 NMR shieldings and named it GIAO-SOS-MP2 method.

As shown in paper **III**, the standard deviations (SD) of the scaled MP2 NMR shieldings are, for almost all basis sets, smaller than those of nonscaled MP2, which was studied for carbon, phosphorus, nitrogen, oxygen, and fluorine NMR shifts. For example, the SD of carbon GIAO-SCS-MP2 shifts using the def2-SVP drops from 6.3 to 2.3 ppm and is also smaller than the CCSD(T) result calculated with the def2-SVP basis (5.7 ppm). For oxygen GIAO-SCS-MP2 shifts, the SD for using the cc-pVDZ basis reduces from 30 to 6.4 ppm in comparison to nonscaled MP2 shifts.

Furthermore, we show the beneficial use of mixed basis sets for the HF-part and the correlation part of carbon MP2 and CCSD(T) NMR shifts. For the GIAO-SCS-MP2 method, e.g., the SD for the basis set 6-31G** drops from 2.4 to 1.7 ppm by calculating the HF part with the qz2p basis. Also, the SD of CCSD(T)/pcS-1 NMR shifts can be improved from 2.1 to 0.9 ppm for the combination corr.-part:pcS-1/ HF-part:pcS-2.

Chapter 3

Conclusion and outlook

In this work, linear- and sublinear-scaling methods for calculating properties at the MP2 level were introduced. The presented methods are based on the AO-MP2 approach, where the rate-determining steps are performed in the local atomic orbital basis. Therefore, these methods open up the pathway for reducing the computational cost of the MP2 method by applying QQR-type integral estimates.

For calculating MP2 NMR shieldings, a method is presented that reduces the computational cost for large molecules from the fifth power to linear, or even sub-linear for selected nuclei. Since the calculations are aimed at large molecules, where often not all the NMR shieldings are of interest, here, the focus is on the sublinear-scaling approach for specific nuclei. The sublinear-scaling behavior can be achieved by avoiding the global magnetic field perturbation in the rate-determining steps, thereby exploiting the local nuclear magnetic spin perturbation. The sublinear-scaling significant elements of the rate-determining steps can be preselected by using our QQR-type estimates for AO-MP2 NMR shieldings, where the differences with respect to MO-MP2 shieldings can be controlled by a screening threshold. Furthermore, our AO-MP2 shieldings can also be included in a quantum-mechanics/molecular-mechanics (QM/MM) approach. Within this approach, the size of the QM region has an influence on the accuracy of the results. For this reason, the convergence of the value of the MP2 NMR shieldings with respect to the size of the QM region needs to be investigated. For this study, our AO-MP2 NMR shieldings are especially well-suited, since very large QM-regions are becoming accessible with our method.

To improve the accuracy of MP2 NMR shieldings, GIAO-SCS-MP2 and GIAO-SOS-MP2 methods are presented, where the opposite and same spin part of the correlation contributions are scaled differently. Thereby, the standard deviations (SD) with respect to CCSD(T)/cc-pVQZ results can be significantly reduced for almost all basis sets in comparison to nonscaled MP2, which is shown for our benchmark sets for carbon, phosphorus, nitrogen, oxygen, and fluorine NMR shifts. These methods can also be extended to investigate other NMR active nuclei.

Furthermore, QQR-type estimates are presented for AO-MP2 gradients, which can be used for geometry optimizations. By applying these estimates the computational cost of the rate-determining steps is reduced from fifth power to linear for sufficiently large molecules, where the deviations between the AO-MP2 and MO-MP2 gradients can be controlled by screening thresholds. Moreover, the AO-MP2 gradients can also be combined with the resolution-of-the-identity (RI) and the Cholesky decomposition of the pseudo-densities (CDD). While the RI-CDD MP2 method is struggling with an asymptotic cubic-scaling, the prefactor can strongly be reduced by using efficient sparse matrix algebra. Therefore, this approach is quite promising to improve the efficiency for calculations with large basis sets.

Bibliography

- [1] A. SZABO and N. OSTLUND, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry Series, Dover Publications, 1996.
- [2] T. HELGAKER, P. JØRGENSEN, and J. OLSEN, *Molecular electronic-structure theory*, Wiley-VCH, Weinheim, 2000.
- [3] C. A. WHITE, B. G. JOHNSON, P. M. W. GILL, and M. HEAD-GORDON, *Chem. Phys. Lett.* **230**, 8 (1994).
- [4] C. A. WHITE, B. G. JOHNSON, P. M. W. GILL, and M. HEAD-GORDON, *Chem. Phys. Lett.* **253**, 268 (1996).
- [5] C. OCHSENFELD, C. A. WHITE, and M. HEAD-GORDON, *J. Chem. Phys.* **109**, 1663 (1998).
- [6] C. OCHSENFELD, *Chem. Phys. Lett.* **327**, 216 (2000).
- [7] R. G. PARR and W. YANG, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1994.
- [8] M. FEYEREISEN, G. FITZGERALD, and A. KOMORNICKI, *Chem. Phys. Lett.* **208**, 359 (1993).
- [9] F. WEIGEND, M. HÄSER, H. PATZELT, and R. AHLRICHS, *Chem. Phys. Lett.* **294**, 143 (1998).
- [10] D. E. BERNHOLDT and R. J. HARRISON, *Chem. Phys. Lett.* **250**, 477 (1996).
- [11] S. SAEBØ, J. BAKER, K. WOLINSKI, and P. PULAY, *J. Chem. Phys.* **120**, 11423 (2004).
- [12] A. EL AZHARY, G. RAUHUT, P. PULAY, and H.-J. WERNER, *J. Chem. Phys.* **108**, 5185 (1998).
- [13] M. SCHÜTZ, H.-J. WERNER, R. LINDH, and F. R. MANBY, *J. Chem. Phys.* **121**, 737 (2004).

- [14] J. ALMLÖF, *Chem. Phys. Lett.* **181**, 319 (1991).
- [15] M. HÄSER and J. ALMLÖF, *J. Chem. Phys.* **96**, 489 (1992).
- [16] M. HÄSER, *Theoret. Chim. Acta* **87**, 147 (1993).
- [17] S. A. MAURER, D. S. LAMBRECHT, D. FLAIG, and C. OCHSENFELD, *J. Chem. Phys.* **136**, 144107 (2012).
- [18] S. A. MAURER, D. S. LAMBRECHT, J. KUSSMANN, and C. OCHSENFELD, *J. Chem. Phys.* **138**, 014101 (2013).
- [19] M. HÄSER and R. AHLRICHS, *J. Comput. Chem.* **10**, 104 (1989).
- [20] J. KUSSMANN and C. OCHSENFELD, *J. Chem. Phys.* **127**, 054103 (2007).
- [21] M. BEER, J. KUSSMANN, and C. OCHSENFELD, *J. Chem. Phys.* **134**, 074102 (2011).
- [22] J. GAUSS and H.-J. WERNER, *Phys. Chem. Chem. Phys.* **2**, 2083 (2000).
- [23] S. LOIBL and M. SCHÜTZ, *J. Chem. Phys.* **137**, 084107 (2012).
- [24] N. C. HANDY and H. F. SCHAEFER III, *J. Chem. Phys.* **81**, 5031 (1984).
- [25] Y. YAMAGUCHI, J. D. GODDARD, Y. OSAMURA, and H. F. SCHAEFER III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press Inc, 1994.
- [26] S. GRIMME, *J. Chem. Phys.* **118**, 9095 (2003).
- [27] S. GRIMME, L. GOERIGK, and R. F. FINK, *WIREs Comput. Mol. Sci.* **2**, 886 (2012).
- [28] Y. JUNG, R. C. LOCHAN, A. D. DUTOI, and M. HEAD-GORDON, *J. Chem. Phys.* **121**, 9793 (2004).
- [29] R. F. FINK, *J. Chem. Phys.* **133**, 174113 (2010).
- [30] S. SCHWEIZER, B. DOSER, and C. OCHSENFELD, *J. Chem. Phys.* **128**, 154101 (2008).
- [31] P. R. SURJÁN, *Chem. Phys. Lett.* **406**, 318 (2005).
- [32] P. Y. AYALA and G. E. SCUSERIA, *J. Chem. Phys.* **110**, 3660 (1999).

- [33] H. HORN, H. WEISS, M. HÄSER, M. EHRLIG, and R. AHLRICHS, *J. Comput. Chem.* **12**, 1058 (1991).
- [34] S. A. MAURER, L. CLIN, and C. OCHSENFELD, *J. Chem. Phys.* **140**, 224112 (2014).
- [35] C. OCHSENFELD and M. HEAD-GORDON, *Chem. Phys. Lett.* **270**, 399 (1997).
- [36] M. BEER and C. OCHSENFELD, *J. Chem. Phys.* **128**, 221102 (2008).

Chapter 5

Publications

- 5.1 Paper I: "A linear- and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Møller-Plesset perturbation theory",
M. Maurer and C. Ochsenfeld,
J. Chem. Phys., 138, 174104 (2013)



A linear- and sublinear-scaling method for calculating NMR shieldings in atomic orbital-based second-order Møller-Plesset perturbation theory

Marina Maurer and Christian Ochsenfeld^{a)}

Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandstr. 7, D-81377 Munich, Germany and Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandstr. 5-13, D-81377 Munich, Germany

(Received 23 January 2013; accepted 26 March 2013; published online 1 May 2013)

An atomic-orbital (AO) based formulation for calculating nuclear magnetic resonance chemical shieldings at the second-order Møller-Plesset perturbation theory level is introduced, which provides a basis for reducing the scaling of the computational effort with the molecular size from the fifth power to linear and for a specific nucleus to sublinear. The latter sublinear scaling in the rate-determining steps becomes possible by avoiding global perturbations with respect to the magnetic field and by solving for quantities that involve the local nuclear magnetic spin perturbation instead. For avoiding the calculation of the second-order perturbed density matrix, we extend our AO-based reformulation of the Z-vector method within a density matrix-based scheme. Our pilot implementation illustrates the fast convergence with respect to the required number of Laplace points and the asymptotic scaling behavior in the rate-determining steps. © 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4801084>]

I. INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is a central tool for gaining insights into structure and dynamics of molecular systems. However, deriving reliable structural information from the experimental spectrum is often a challenging task, where quantum-chemical calculations can provide useful complementary insights for the interpretation. Here, already the most simple quantum-chemical approximations such as Hartree-Fock (HF)¹⁻⁴ and density-functional theory (DFT)⁵⁻⁷ can provide often very useful NMR chemical shieldings. However, for more reliable and most accurate results, the use of higher level wavefunction-based methods such as second-order Møller-Plesset perturbation theory (MP2),⁸⁻¹⁰ coupled-cluster (CC),¹¹⁻¹³ or multiconfigurational self-consistent-field (MCSCF)^{14,15} methods is highly desirable. However, with the increased level of accuracy and reliability, the computational effort increases dramatically and applications become restricted to rather small molecules. Nevertheless, already MP2 theory has proven to provide often highly accurate relative NMR chemical shieldings,^{9,10,16-19} while absolute NMR shieldings are of less importance in comparing with experimental data. Furthermore, it needs mentioning that for all methods using finite basis-set representations for computing NMR shieldings, the gauge-origin problem is of central importance.^{1,20-23} Here, the use of so-called gauge-including atomic orbitals (GIAOs)^{1,20,24} has proven to be the most successful approach,^{2,23} which is also used throughout our present work.

While over the last decade linear- and sublinear-scaling HF and DFT methods for calculating NMR shieldings have been developed that allow for computing molecular sys-

tems with more than 1000 atoms on simple workstation computers,²⁵⁻²⁸ the computational scaling even for the simplest wavefunction-based correlation method, MP2 theory, increases as $\mathcal{O}(N^5)$ with the system size N hampering its application to larger molecular systems. The theory for calculating NMR shieldings at the molecular-orbital (MO) MP2 level was introduced first by Gauss in 1992,^{8,10} who also presented MO-based Z-vector equations for this context (see also related work for dipoles²⁹).

As mentioned above, the main limitation of the MO-based approach is the scaling behavior of the central processing unit (CPU) time due to the transformation of the perturbed and unperturbed four-center two-electron integrals to the MO representation ($\mathcal{O}(N^5)$ scaling), but also due to the storage requirements for the perturbed and unperturbed two-electron integrals scaling with $\mathcal{O}(N^4)$. Therefore, the MO-MP2 method is so far limited to small molecules. In order to reduce the scaling behavior of the storage requirements from $\mathcal{O}(N^4)$ to $\mathcal{O}(N^3)$, Kollwitz *et al.*^{30,31} introduced an integral-direct GIAO-MP2 implementation and by exploiting molecular point group symmetry the CPU time was decreased. Furthermore, a local-correlation approximation was introduced for reducing the scaling behavior by Gauss and Werner³² and most recently the efficiency was improved by a combination with an additional density fitting scheme³³ by Loibl and Schütz. However, to the best of our knowledge no linear scaling for the calculation of MP2 NMR shieldings has been demonstrated in the literature. The same holds for sublinear scaling of course.

In our present work, we derive an AO-based formulation of NMR shieldings at the MP2 level, that opens the way for calculating MP2 NMR shieldings of large molecules with an effort scaling linearly or even sublinearly with system size in the rate-determining steps. The latter sublinear scaling is

^{a)}Electronic mail: christian.ochsenfeld@uni-muenchen.de

the extension of an earlier formulation for nuclei-selected NMR shieldings at the HF and DFT levels introduced by Beer *et al.*²⁸ Here, the shielding tensor for just a few nuclei of interest for a large molecule or for a molecular system within a solvent environment is computed. As starting point for our present AO-MP2 NMR reformulation, we employ the Laplace approach originally introduced by Almlöf and Häser^{34–36} for avoiding the orbital energy denominator in calculating MP2 energies and energy gradients. This Laplace approach was employed earlier by us for devising a linear-scaling AO-MP2 method by preselection of numerically significant contributions using distance-including integral estimates denoted as QQR.^{37,38} Using this preselection method, the largest molecular system computed so far at the AO-MP2 level was a DNA repair complex with 2025 atoms and 20 371 basis functions.³⁸

We start by presenting an AO-based reformulation of the MP2 NMR chemical shielding tensor opening the way to exploit locality in order to allow for linear scaling. For efficient calculations, we present an AO-based reformulation of the Z-vector method of Handy and Schaefer^{39,40} within a density matrix-based scheme to avoid the second derivative of the density matrix. In a second step, we show, how to avoid global perturbations with respect to the magnetic field and employ local quantities perturbed with respect to the nuclear magnetic moment in order to allow for sublinear scaling for a specific nucleus.

Therefore, we exploit the AO-based Z-vector technique developed by Schweizer *et al.*⁴¹ to avoid the linear-scaling determination of the perturbed density matrix with respect to the magnetic field. An alternative pathway for a reformulation could be a Lagrangian approach (for a description of an MO-based formulation see, e.g., Ref. 23), which we have not used in our present work. Using our new AO-based reformulation of MP2 shieldings, we obtained a first pilot implementation for which we present first results with respect to the number of required Laplace points and the asymptotic scaling behavior of intermediate quantities. These data illustrate the possibility for an asymptotic sublinear scaling in the rate-determining steps.

Our formulation presented here focuses on calculating only a few selected nuclei with sublinear scaling effort. While this allows, in principle, a linear-scaling calculation of shieldings for all nuclei, in such a case a modified, closely related, formulation would be more efficient, which will be presented in future work.

II. THEORY

A. AO-MP2 energy expression

As starting point for deriving a fully AO-based method for the calculation of MP2 NMR chemical shieldings suitable for linear or sublinear scaling, we employ the Laplace-based energy formulation originally introduced by Almlöf and Häser:^{34–36}

$$E_{\text{AO-MP2}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} E_{JK}^{(\alpha)} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma}) \times [2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)], \quad (1)$$

with the fully transformed integrals:

$$(\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma}) = \sum_{\mu'} \underline{P}_{\mu'\mu} \left(\sum_{\nu'} \bar{P}_{\nu'\nu} \left(\sum_{\lambda'} \underline{P}_{\lambda'\lambda} \left(\sum_{\sigma'} \bar{P}_{\sigma'\sigma} (\mu'\nu'|\lambda'\sigma') \right) \right) \right), \quad (2)$$

and the occupied and unoccupied pseudo-density matrices:

$$\underline{P}_{\mu'\mu} = \sum_i C_{\mu'i} e^{\varepsilon_i t_a} C_{\mu i} = (e^{t_a \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}})_{\mu'\mu}, \quad (3)$$

$$\bar{P}_{\nu'\nu} = \sum_a C_{\nu'a} e^{-\varepsilon_a t_a} C_{\nu a} = (e^{-t_a \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}})_{\nu'\nu}.$$

Here, \mathbf{P}_{occ} corresponds to the conventional one-particle density matrix (summation over the occupied space), whereas \mathbf{P}_{virt} denotes the virtual density matrix (summation of MOs of the virtual space). The latter equation shows that also the formation of the pseudo-densities can be done without any use of canonical MOs and entirely in a density matrix-based fashion.^{41–43} By using our most recently introduced distance-including two-electron integral estimates denoted by QQR,^{37,38} it is possible to preselect numerically significant integral products in Eq. (1), such that the MP2 energy can be calculated in a linear-scaling fashion for molecules with a significant HOMO-LUMO gap and molecules with more than 1000 atoms and 10 000 basis functions become accessible on a single core computer.

B. Laplace-based AO-MP2 chemical shielding tensor

The magnetic shielding tensor is obtained by the second derivatives of the energy with respect to the magnetic field \mathbf{B} and the nuclear magnetic spin moment \mathbf{m}_j :

$$\sigma_j = \left(\frac{\partial^2 E}{\partial \mathbf{m}_j \partial \mathbf{B}} \right)_{\mathbf{B}, \mathbf{m}_j=0}. \quad (4)$$

In order to obtain a suitable AO-based reformulation for the chemical shielding tensor, we employ the AO-MP2 energy expression (Eq. (1)) and differentiate first with respect to the magnetic field \mathbf{B} and then to the nuclear magnetic spin moment \mathbf{m}_j . We have chosen this order of differentiation, since in the following we will substitute the linear scaling density matrix with respect to the magnetic field \mathbf{B} by a sublinear scaling term perturbed with respect to the nuclear magnetic moment. The GIAO-approach^{1,20,23,24} is employed throughout in order to avoid the gauge-origin problem by using B-field dependent basis functions.

1. The first derivative with respect to the magnetic field \mathbf{B}

The first derivative of the AO-MP2 energy with respect to the magnetic field \mathbf{B} is similar to the AO-MP2 gradients introduced by Schweizer *et al.* in 2008⁴¹ (an explicit derivation of the following expression as well as the definition of $\bar{\mathbf{Y}}_1$, \mathbf{Y}_1 ,

$\bar{\mathbf{Y}}_2$ and \mathbf{Y}_2 is shown in Appendix A):

$$\begin{aligned} E_{\text{AO-MP2}}^{\mathbf{B}} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2\mathcal{I}^{\mathbf{B}} \\ & + 2\text{Tr}[(\bar{\mathbf{Y}}_1 - \mathbf{Y}_1 + \mathbf{G}[\bar{\mathbf{Y}}_2 + \mathbf{Y}_2] \\ & + \bar{\mathbf{R}}e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} - \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})\mathbf{P}_{\text{occ}}^{\mathbf{B}}] \\ & + 2\text{Tr}[(\bar{\mathbf{Y}}_2 + \mathbf{Y}_2)\mathbf{F}^{(\mathbf{B})}] \\ & + 2\text{Tr}[-(\mathbf{Y}_1 + \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1}] \} \\ = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2\mathcal{I}^{\mathbf{B}} + 2\text{Tr}[\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{B}}] \\ & + 2\text{Tr}[\mathcal{F}\mathbf{F}^{(\mathbf{B})}] + 2\text{Tr}[\mathcal{S}\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1}] \}, \quad (5) \end{aligned}$$

with

$$\mathcal{P} = \bar{\mathbf{Y}}_1 - \mathbf{Y}_1 + \mathbf{G}[\bar{\mathbf{Y}}_2 + \mathbf{Y}_2] + \bar{\mathbf{R}}e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} - \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}}, \quad (6a)$$

$$\mathcal{F} = \bar{\mathbf{Y}}_2 + \mathbf{Y}_2, \quad (6b)$$

$$\mathcal{S} = -(\mathbf{Y}_1 + \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}}) \quad (6c)$$

and

$$\begin{aligned} \mathcal{I}^{\mathbf{B}} = & \sum_{\mu\nu\lambda\sigma} (\underline{\mu}|\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}}, \\ \bar{R}_{\mu'\mu} = & \sum_{\nu\lambda\sigma} (\underline{\mu}'\bar{\nu}|\underline{\lambda}\bar{\sigma})[2(\mu\nu||\lambda\sigma) - (\mu\sigma||\lambda\nu)], \quad (7) \\ \underline{R}_{\nu'\nu} = & \sum_{\mu\lambda\sigma} (\underline{\mu}\bar{\nu}'|\underline{\lambda}\bar{\sigma})[2(\mu\nu||\lambda\sigma) - (\mu\sigma||\lambda\nu)]. \end{aligned}$$

To formulate the AO-MP2 shielding tensor, this expression has to be furthermore differentiated with respect to the nuclear magnetic moment (to be shown in Sec. II B 2).

2. Second derivatives by further differentiation: The derivative with respect to the nuclear magnetic spin moment \mathbf{m}_j

To obtain the expression for the chemical shielding tensor one has to differentiate the gradient in Eq. (5) with respect to the nuclear magnetic spin moment \mathbf{m}_j leading to:

$$\begin{aligned} \sigma_j^{MP2} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \frac{d^2 E_{JK}^{(\alpha)}}{d\mathbf{m}_j d\mathbf{B}} \\ = & - \sum_{\alpha=1}^{\tau} w_{\alpha} [2(\mathcal{I}^{\mathbf{B}})^{\mathbf{m}_j} + 2\text{Tr}(\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{B}})^{\mathbf{m}_j} + 2\text{Tr}(\mathcal{F}\mathbf{F}^{(\mathbf{B})})^{\mathbf{m}_j} \\ & + 2\text{Tr}(\mathcal{S}\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1})^{\mathbf{m}_j}]. \quad (8) \end{aligned}$$

Here, the first term is given by

$$\begin{aligned} (\mathcal{I}^{\mathbf{B}})^{\mathbf{m}_j} = & 2 \sum_{\mu\nu\lambda\sigma} \sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\underline{\mu}'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}} \\ & + 2 \sum_{\mu\nu\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\underline{\mu}\bar{\nu}'|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}}. \quad (9) \end{aligned}$$

The second term in Eq. (8) is

$$(\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{B}})^{\mathbf{m}_j} = \mathcal{P}^{\mathbf{m}_j}\mathbf{P}_{\text{occ}}^{\mathbf{B}} + \mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}, \quad (10)$$

with

$$\begin{aligned} \mathcal{P}^{\mathbf{m}_j}\mathbf{P}_{\text{occ}}^{\mathbf{B}} = & [\bar{\mathbf{Y}}_1^{\mathbf{m}_j} - \mathbf{Y}_1^{\mathbf{m}_j} + \mathbf{G}[\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \mathbf{Y}_2^{\mathbf{m}_j}] + \bar{\mathbf{R}}e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} \\ & + \bar{\mathbf{R}}(e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}})^{\mathbf{m}_j} - \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} \\ & - \mathbf{R}(e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j}]\mathbf{P}_{\text{occ}}^{\mathbf{B}}. \quad (11) \end{aligned}$$

The third term in Eq. (8) reads as

$$(\mathcal{F}\mathbf{F}^{(\mathbf{B})})^{\mathbf{m}_j} = (\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \mathbf{Y}_2^{\mathbf{m}_j})\mathbf{F}^{(\mathbf{B})} + (\bar{\mathbf{Y}}_2 + \mathbf{Y}_2)\mathbf{F}^{(\mathbf{Bm}_j)}, \quad (12)$$

with the second derivative of the Fock matrix given by

$$\mathbf{F}^{(\mathbf{Bm}_j)} = \mathbf{h}^{\mathbf{Bm}_j} + \mathbf{G}^{\mathbf{B}}[\mathbf{P}_{\text{occ}}^{\mathbf{m}_j}]. \quad (13)$$

Finally, the last term of Eq. (8) is

$$\begin{aligned} (\mathcal{S}\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1})^{\mathbf{m}_j} = & [- (\mathbf{Y}_1^{\mathbf{m}_j} + \mathbf{R}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} + \mathbf{R}(e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j})] \mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1}. \quad (14) \end{aligned}$$

It is worthwhile to mention that the basis functions are independent of the perturbation \mathbf{m}_j , so that also the overlap matrix is independent and $\mathbf{S}^{\mathbf{m}_j} = \mathbf{0}$.

The derivatives of the recursion formulae $\bar{\mathbf{Y}}_1$, \mathbf{Y}_1 , $\bar{\mathbf{Y}}_2$, and \mathbf{Y}_2 occurring in Eqs. (11), (12), and (14) are given in Appendix B. Moreover, for the terms in Eq. (9) the derivative of the pseudo-densities with respect to the nuclear magnetic spin moment \mathbf{m}_j is required

$$\begin{aligned} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} = & (e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}})^{\mathbf{m}_j} \mathbf{P}_{\text{occ}} + e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} \mathbf{P}_{\text{occ}}^{\mathbf{m}_j}, \\ \bar{P}_{\nu'\nu}^{\mathbf{m}_j} = & (e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j} \mathbf{P}_{\text{virt}} + e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} \mathbf{P}_{\text{virt}}^{\mathbf{m}_j}. \quad (15) \end{aligned}$$

Here, the derivatives of the exponentials are needed, which can be calculated via recursion $\bar{\mathbf{T}} = (e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}})^{\mathbf{m}_j} = \sum_{k=1}^m \bar{\mathbf{T}}^{(k)}$, in which each term is obtained by

$$\begin{aligned} \bar{\mathbf{T}}^{(k)} = & \frac{1}{k} [\mathbf{T}^{(k-1)}(t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}) \\ & + (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(k-1)}(t_{\alpha}\mathbf{P}_{\text{occ}}^{\mathbf{m}_j}\mathbf{F} + t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}^{\mathbf{m}_j})]. \quad (16) \end{aligned}$$

The same applies to $(e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j}$, except one uses the virtual density matrix. Note that $\mathbf{P}_{\text{virt}}^{\mathbf{m}_j}$ can be expressed in terms of $\mathbf{P}_{\text{occ}}^{\mathbf{m}_j}$ as shown in Appendix B.

Furthermore, the perturbed pseudo-densities are required for the derivative of the R-matrices:

$$\begin{aligned} \underline{R}_{\nu'\nu}^{\mathbf{m}_j} = & \sum_{\mu'\mu} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\underline{\mu}'\bar{\nu}'|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma) \\ & + \sum_{\lambda'\lambda} \underline{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\underline{\mu}\bar{\nu}'|\underline{\lambda}'\bar{\sigma})(\mu\nu||\lambda\sigma) \\ & + \sum_{\sigma'\sigma} \bar{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\underline{\mu}\bar{\nu}'|\underline{\lambda}\bar{\sigma}')(\mu\nu||\lambda\sigma), \\ \bar{R}_{\mu'\mu}^{\mathbf{m}_j} = & \sum_{\nu'\nu} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\underline{\mu}'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma) \\ & + \sum_{\lambda'\lambda} \bar{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\underline{\mu}'\bar{\nu}|\underline{\lambda}'\bar{\sigma})(\mu\nu||\lambda\sigma) \\ & + \sum_{\sigma'\sigma} \bar{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\underline{\mu}'\bar{\nu}|\underline{\lambda}\bar{\sigma}')(\mu\nu||\lambda\sigma). \quad (17) \end{aligned}$$

Overall, the AO-based MP2 shielding tensor is given by

$$\begin{aligned} \sigma_j^{MP2} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ 4 \sum_{\mu\nu\lambda\sigma} \sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}} \right. \\ & + 4 \sum_{\mu\nu\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}} \\ & + 2\text{Tr}[(\bar{\mathbf{Y}}_1^{\mathbf{m}_j} - \mathbf{Y}_1^{\mathbf{m}_j} + \mathbf{G}[\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \mathbf{Y}_2^{\mathbf{m}_j}] + \bar{\mathbf{R}}^{\mathbf{m}_j} e^{t_{\alpha}} \mathbf{P}_{\text{occ}}^{\mathbf{F}} \\ & + \bar{\mathbf{R}}(e^{t_{\alpha}} \mathbf{P}_{\text{occ}}^{\mathbf{F}})^{\mathbf{m}_j} - \mathbf{R}^{\mathbf{m}_j} e^{-t_{\alpha}} \mathbf{P}_{\text{virt}}^{\mathbf{F}} - \mathbf{R}(e^{-t_{\alpha}} \mathbf{P}_{\text{virt}}^{\mathbf{F}})^{\mathbf{m}_j}) \mathbf{P}_{\text{occ}}^{\mathbf{B}}] \\ & + 2\text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}] \\ & + 2\text{Tr}[(\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \mathbf{Y}_2^{\mathbf{m}_j}) \mathbf{F}^{(\mathbf{B})} + (\bar{\mathbf{Y}}_2 + \mathbf{Y}_2) \mathbf{F}^{(\mathbf{Bm}_j)}] \\ & + 2\text{Tr}[(- (\mathbf{Y}_1^{\mathbf{m}_j} + \mathbf{R}^{\mathbf{m}_j} e^{-t_{\alpha}} \mathbf{P}_{\text{virt}}^{\mathbf{F}} \\ & \left. + \mathbf{R}(e^{-t_{\alpha}} \mathbf{P}_{\text{virt}}^{\mathbf{F}})^{\mathbf{m}_j}) \mathbf{S}^{-1} \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1}] \right\}. \quad (18) \end{aligned}$$

While this equation already represents a fully AO-based reformulation of the NMR chemical shielding tensor and in principle allows for a linear-scaling calculation of the NMR chemical shielding tensor, the second derivative of the one-particle density matrix needs to be avoided as outlined in Subsection II B 3.

3. AO-based Z-vector method for avoiding the second-order density matrix

In order to avoid computation of the second derivatives of the one-particle density matrix $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$, we employ and extend the Z-vector method originally introduced in the MO basis by Handy and Schaefer,^{39,40} or, more specifically its AO-based reformulation introduced by Schweizer *et al.*⁴¹

To allow for application of the Z-vector method, the equations need to be in the following form:

$$E_{\text{AO-MP2}}^{\mathbf{Bm}_j} = \text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}] + \text{Tr}[\mathcal{X}], \quad (19)$$

where the first term depends on $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$, while the second does not. Equation (18) derived above has already been brought into this suitable form.

For a suitable formulation of the Z-vector equations in the AO-basis, we start from the density matrix-based coupled-perturbed self-consistent field equations (D-CPSCF)^{26,44} which can be abbreviated as (we denote \mathbf{P}_{occ} by \mathbf{P} in the following for simplicity and the explicit definitions of the D-CPSCF equations are given in Appendix C):

$$\underline{\mathbf{A}} \mathbf{P}^{\mathbf{B}} = \mathbf{b}^{\mathbf{B}}. \quad (20)$$

Differentiation with respect to \mathbf{m}_j leads to

$$\underline{\mathbf{A}} \mathbf{P}^{\mathbf{Bm}_j} + \underline{\mathbf{A}}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}} = \mathbf{b}^{\mathbf{Bm}_j}. \quad (21)$$

Here, $\underline{\mathbf{A}}$ is the symmetric, positive-definite Hessian that can in principle be inverted. By multiplying from the left with the inverse of $\underline{\mathbf{A}}$, one obtains

$$\mathbf{P}^{\mathbf{Bm}_j} = \underline{\mathbf{A}}^{-1} \mathbf{b}^{\mathbf{Bm}_j} - \underline{\mathbf{A}}^{-1} \underline{\mathbf{A}}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}}. \quad (22)$$

Multiplication with \mathcal{P} from the left leads to the definition of the Z-vector:

$$\begin{aligned} \mathcal{P} \mathbf{P}^{\mathbf{Bm}_j} &= \mathcal{P} \underline{\mathbf{A}}^{-1} (\mathbf{b}^{\mathbf{Bm}_j} - \underline{\mathbf{A}}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}}), \\ &= \underbrace{\mathcal{P} \underline{\mathbf{A}}^{-1}}_{=\mathbf{Z}^T} \mathbf{b}^{\mathbf{Bm}_j} - \mathbf{Z}^T \underline{\mathbf{A}}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}}, \end{aligned} \quad (23)$$

so that the term $\text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}]$ in Eq. (18) can be substituted by the much simpler term:

$$\text{Tr}[\mathcal{P} \mathbf{P}^{\mathbf{Bm}_j}] = \text{Tr}[\mathbf{Z}^T (\mathbf{b}^{\mathbf{Bm}_j} - \underline{\mathbf{A}}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}})]. \quad (24)$$

In our implementation, we employ the most efficient Laplace-based D-CPSCF equations (DL-CPSCF) as introduced by Beer and Ochsenfeld.²⁷ The explicit equations to be solved are shown in Appendix D. The major advantage of the Z-vector method is that the iterative determination of the Z-vector does not depend upon the perturbation and thereby only one equation has to be solved, in contrast to the conventional $9N_{\text{atoms}}$ perturbations (and the equivalent number of CPSCF equations) if $\mathbf{P}^{\mathbf{Bm}_j}$ would have been necessary.

C. Nuclei-selected AO-MP2 shieldings—A sublinear scaling reformulation

In a large molecular system or in a system within an explicit solvent environment, typically just a few NMR chemical shieldings are of interest. Therefore, we aim to derive a nuclei-selected formulation of Laplace-based AO-MP2 chemical shieldings that allows to achieve sublinear scaling in the rate-determining steps. For the formulation we follow the path of the nuclei-selected NMR formulation originally introduced at HF and DFT levels by Beer *et al.*²⁸

In our AO-based formulation of the MP2 shielding tensor presented so far, the perturbed density matrix with respect to the magnetic field \mathbf{B} occurs. Since \mathbf{B} is a non-local perturbation and the corresponding perturbed matrices $\mathbf{P}^{\mathbf{B}}$ scale linearly with system size, such perturbed quantities need to be avoided and the equations reformulated in terms of the nuclear magnetic moment \mathbf{m}_j which is only a local perturbation (such that the number of numerically significant elements in the corresponding perturbed matrices scales sublinearly, i.e., independent of system size in the asymptotic regime)—for details at the HF and DFT levels, see Ref. 28.

In order to avoid $\mathbf{P}^{\mathbf{B}}$, we employ the AO-based Z-vector equations. For a suitable formulation of each term containing the perturbed density matrix with respect to the magnetic field \mathbf{B} , cyclic permutations are performed within the trace such that the perturbed density matrix is brought to the far right-hand side to allow for application of the Z-vector method. This comprises Eq. (11) and the Z-vector equations derived for avoiding the second derivative of the density matrix (D8), (D3a), and (D3b). All these terms are collected in $\mathcal{O}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}}$ (explicit definitions are given in Appendix E).

To obtain suitable Z-vector equations for these terms, we start with the abbreviated D-CPSCF equations (for details of the DL-CPSCF method employed see discussion in Appendix E)

$$\underline{\mathbf{A}} \mathbf{P}^{\mathbf{B}} = \mathbf{b}^{\mathbf{B}}, \quad (25)$$

TABLE I. Correlation contribution to ^{13}C -NMR shieldings (in ppm) at MP2 level in dependence of the number of Laplace points. In addition, the difference to the MO-MP2 shieldings is given in parentheses (6-311G** basis for CH_4 , CHCH , and CH_3CH_3 , and 3-21G for the other systems).

Molecule	Number of Laplace points				
	1	2	3	4	5
CH_4	4.7 (1.3)	5.9 (0.0)	5.9 (0.0)	5.9 (0.0)	5.9 (0.0)
CHCH	6.2 (4.6)	9.6 (1.2)	10.6 (0.2)	10.7 (0.1)	10.8 (0.0)
CH_3CH_3	3.6 (0.5)	4.2 (−0.1)	4.1 (0.0)	4.1 (0.0)	4.1 (0.0)
CH_2CH_2	17.3 (4.1)	20.4 (1.0)	21.2 (0.2)	21.4 (0.0)	21.4 (0.0)
$\text{CH}_2(\text{CH})_2\text{CH}_2$	15.1 (4.0)	18.1 (1.0)	18.8 (0.3)	19.0 (0.1)	19.1 (0.0)
$\text{CH}_2(\underline{\text{CH}})_2\text{CH}_2$	15.0 (2.8)	17.0 (0.8)	17.7 (0.2)	17.8 (0.0)	17.9 (0.0)
$\text{CH}_3(\text{CH})_2\text{CH}_3$	3.7 (−0.3)	3.6 (−0.1)	3.5 (0.0)	3.5 (0.0)	3.5 (0.0)
$\text{CH}_3(\underline{\text{CH}})_2\text{CH}_3$	3.0 (−0.7)	2.5 (−0.2)	2.3 (0.0)	2.3 (0.0)	2.3 (0.0)

multiply from the left with $\underline{\mathbf{A}}^{-1}$

$$\mathbf{P}^{\mathbf{B}} = \underline{\mathbf{A}}^{-1} \mathbf{b}^{\mathbf{B}}, \quad (26)$$

with $\mathcal{O}^{\mathbf{m}_j}$, so that the Z-vector results as

$$\begin{aligned} \mathcal{O}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}} &= \mathcal{O}^{\mathbf{m}_j} \underline{\mathbf{A}}^{-1} \mathbf{b}^{\mathbf{B}}. \\ &= \underbrace{(\mathbf{Z}^T)^{\mathbf{m}_j}} \end{aligned} \quad (27)$$

In this way, the term $\text{Tr}[\mathcal{O}^{\mathbf{m}_j} \mathbf{P}^{\mathbf{B}}]$ containing the perturbed density matrix with respect to the magnetic field can be substituted by $\text{Tr}[(\mathbf{Z}^T)^{\mathbf{m}_j} \mathbf{b}^{\mathbf{B}}]$. The major advantage of this formulation is the sublinear scaling ($\mathcal{O}(N^0)$) for a specific nucleus in the iterative solution for $(\mathbf{Z}^T)^{\mathbf{m}_j}$, that scales sublinearly in contrast to the linear scaling determination of $\mathbf{P}^{\mathbf{B}}$.

III. FIRST PRELIMINARY IMPLEMENTATION AND SCALING BEHAVIOR

At this stage of method development, we have implemented the AO-MP2 shielding tensor formulation described above in an inefficient way in order to support the validity of all equations. Therefore, any applications to molecular systems are presently constrained to small molecules and basis sets. In Secs. III A–III C, we will briefly discuss first results with respect to the required number of Laplace points and the expected scaling behavior determined by counting the number of numerically significant elements in the occurring terms of the formulation.

A. Required number of Laplace points

First, the number of Laplace points necessary for reliable results is investigated. For this purpose, the structures of the test set used by Flaig *et al.*¹⁸ are employed for studying the accuracy of the isotropic NMR shieldings. The corresponding results are summarised in Table I for ^{13}C -NMR shieldings of several alkanes, alkenes, and alkynes employing different basis sets. Here, both the MP2-contribution (without the HF-part) as well as the difference to MO-MP2 shieldings (the latter were calculated with the TURBOMOLE program package⁴⁵) are listed. The results show that typically three to four Laplace points are sufficient to reduce the error to 0.1 ppm at most. Similarly, for the ^1H -NMR shieldings listed in Table II three Laplace points provide sufficient accuracy. These results indicate that just a few Laplace points seem necessary for reliable MP2 NMR shieldings, which is in line with the observations made for the calculation of AO-MP2 energies and gradients, while NMR shieldings seem even less sensitive.

B. Locality of MP2 shieldings and comparison to HF shieldings

As shown by Beer *et al.*,²⁸ it is possible to calculate the HF- or also DFT-based shieldings for a specific nucleus with a sublinear-scaling computational effort. An example for the local behavior is illustrated in Fig. 1 which plots the difference

TABLE II. Correlation contribution to ^1H -NMR shieldings (in ppm) at MP2 level in dependence of the number of Laplace points. In addition, the difference to the MO-MP2 shieldings is given in parentheses (6-311G** basis for CH_4 , CHCH , and CH_3CH_3 , and 3-21G for the other systems).

Molecule	Number of Laplace points				
	1	2	3	4	5
CH_4	−0.03 (−0.06)	−0.07 (−0.02)	−0.09 (0.00)	−0.09 (0.00)	−0.09 (0.00)
CHCH	−0.03 (−0.08)	−0.10 (−0.01)	−0.11 (0.00)	−0.11 (0.00)	−0.11 (0.00)
CH_3CH_3	−0.12 (−0.13)	−0.22 (−0.03)	−0.24 (−0.01)	−0.25 (0.00)	−0.25 (0.00)
CH_2CH_2	0.28 (0.07)	0.31 (0.04)	0.34 (0.01)	0.35 (0.00)	0.35 (0.00)
$\text{CH}_2(\text{CH})_2\text{CH}_2$ (<i>cis</i>)	0.15 (0.03)	0.14 (0.03)	0.17 (0.01)	0.17 (0.00)	0.18 (0.00)
$\text{CH}_2(\text{CH})_2\text{CH}_2$ (<i>trans</i>)	0.18 (0.03)	0.18 (0.04)	0.20 (0.01)	0.21 (0.00)	0.21 (0.00)
$\text{CH}_2(\underline{\text{CH}})_2\text{CH}_2$	0.27 (0.05)	0.27 (0.04)	0.30 (0.01)	0.31 (0.00)	0.32 (0.00)
$\text{CH}_3(\text{CH})_2\text{CH}_3$	−0.32 (−0.18)	−0.48 (−0.02)	−0.50 (0.00)	−0.50 (0.00)	−0.50 (0.00)
$\text{CH}_3(\underline{\text{CH}})_2\text{CH}_3$	−0.36 (−0.18)	−0.53 (−0.02)	−0.55 (0.00)	−0.55 (0.00)	−0.55 (0.00)

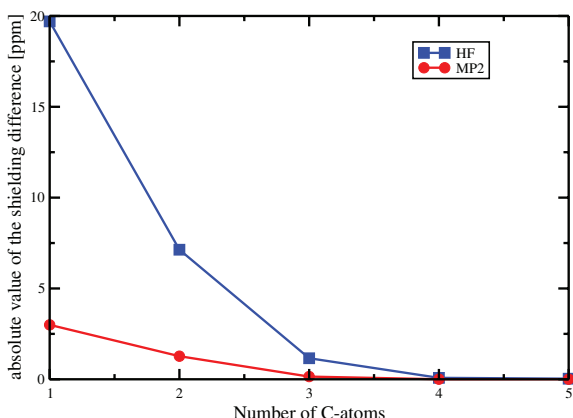


FIG. 1. Absolute value of ¹³C-NMR shielding difference (in ppm) between the first carbon atom of various alkanes and the first carbon in C₆H₁₄ at HF level and the correlation contribution at MP2 level (cc-pVTZ basis; MP2 results were calculated with the TURBOMOLE program package⁴⁵).

of the ¹³C-NMR shielding for the first carbon atom of a variety of simple alkane chains as compared to the shielding of the first carbon in C₆H₁₄. The plot shows the quick decrease of the influence of a lengthening of the alkane chain and fast convergence of the carbon shielding. This behavior is observed for both HF and MP2, while the MP2 convergence seems even faster. Therefore, the sublinear-scaling behavior could be expected for even smaller molecules at the MP2 level.

C. Expected scaling behavior for calculating the AO-MP2 shielding tensor

The most time-consuming steps for building the AO-MP2 shielding tensor are the transformations of the two-electron integrals and most of the contractions. In addition, the contractions of the two-electron integrals with the Z-vector for the first and second derivative are important. Therefore, we focus on these steps in the following and study the expected scaling behavior by counting the numerically significant elements for a series of linear alkanes. The scaling behavior $\mathcal{O}(N^x)$ is determined with respect to the next larger molecule, where the system size N is measured by the number of basis functions. As mentioned above, our pilot code implementation is very preliminary and inefficient, so that we are forced to constrain ourselves to the small STO-3G basis in order to access also larger alkanes. Since we only aim to show the principle scaling behavior that can be reached in the long run, we expect the results obtained by the chosen simple alkane chains and basis sets to be representative also for other molecules with a significant HOMO-LUMO gap and larger basis sets. Here, we do not focus on the onset of the reduced scaling, but rather the possible asymptotic scaling.

1. Scaling behavior of transformations with perturbed pseudo-densities

The results for the transformation steps with the perturbed pseudo-densities and the subsequent contractions, presented in Eqs. (9) and (17), are listed in Table III. All of these transformation steps have a formal scaling behavior of $\mathcal{O}(N^5)$ for each of the $3 \times N_{\text{atoms}}$ perturbations \mathbf{m}_j . The advantage of the nuclei-selected formulation comes into play if the NMR

TABLE III. Number of significant elements and scaling behavior (in parentheses) of transformations and contractions involving the perturbed pseudo-densities as shown in Eqs. (9) and (17) (for calculating ¹³C-NMR shieldings of the first and sixth C-atom of C₁₅H₃₂ or C₂₀H₄₂, respectively; one Laplace point; $\vartheta = 10^{-6}$; basis set STO-3G; perturbations $\mathbf{m}_{j,x}$ and \mathbf{B}_x). For comparison also the scaling behavior of the perturbed and unperturbed pseudo-densities is listed.

	C ₁			C ₆		
	C ₁₅ H ₃₂	C ₂₀ H ₄₂		C ₁₅ H ₃₂	C ₂₀ H ₄₂	
$\underline{P}^{\mathbf{m}_j}$	4540	5578	(0.7)	5924	7474	(0.8)
$\overline{P}^{\mathbf{m}_j}$	4530	5556	(0.7)	5950	7498	(0.8)
\underline{P}	9195	14 542	(1.6)	9195	14 542	(1.6)
\overline{P}	9107	14 414	(1.6)	9107	14 414	(1.6)
$\sum_{v'} \overline{P}_{v'v}^{\mathbf{m}_j} (\mu'v' \underline{\lambda}\overline{\sigma})$	14 170 681	14 787 794	(0.2)	32 069 721	33 646 251	(0.2)
$\sum_{\sigma'} \overline{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\mu'v' \underline{\lambda}\sigma')$	16 885 794	19 299 083	(0.5)	35 910 413	40 846 090	(0.5)
$\sum_{\lambda'} \underline{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\mu'v' \lambda'\overline{\sigma})$	16 399 958	18 613 155	(0.4)	34 966 908	39 499 744	(0.4)
$\overline{\mathbf{R}}^{\mathbf{m}_j}$ (contraction)	3 179 727	3 198 572	(0.0)	6 655 129	6 705 240	(0.0)
$\sum_{\sigma'} \overline{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\underline{\mu}v' \underline{\lambda}\sigma')$	17 370 426	19 917 597	(0.5)	36 833 669	42 076 861	(0.5)
$\sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'v' \underline{\lambda}\overline{\sigma})$	13 751 479	14 314 778	(0.1)	30 880 105	32 318 446	(0.2)
$\sum_{\lambda'} \underline{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\underline{\mu}v' \lambda'\overline{\sigma})$	16 329 701	18 659 577	(0.5)	34 664 724	39 456 424	(0.5)
$\underline{\mathbf{R}}^{\mathbf{m}_j}$ (contraction)	3 077 713	3 091 493	(0.0)	6 481 792	6 522 778	(0.0)
$\sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'v' \underline{\lambda}\overline{\sigma})$	10 470 407	10 763 022	(0.1)	24 015 606	24 739 096	(0.1)
$(\underline{\mu}v' \underline{\lambda}\overline{\sigma}) (\underline{\mu}v' \underline{\lambda}\sigma)^{\mathbf{B}}$	203 529	222 690	(0.3)	432 639	457 835	(0.2)
$\sum_{v'} \overline{P}_{v'v}^{\mathbf{m}_j} (\underline{\mu}v' \underline{\lambda}\overline{\sigma})$	10 420 546	10 746 010	(0.1)	24 039 344	24 837 893	(0.1)
$(\underline{\mu}v' \underline{\lambda}\overline{\sigma}) (\underline{\mu}v' \underline{\lambda}\sigma)^{\mathbf{B}}$	162 851	176 850	(0.3)	351 447	369 837	(0.2)

shielding tensor is of interest for just a few nuclei, where the scaling can be reduced to sublinear in the rate-determining steps.

In order to explore the possible asymptotic scaling behavior in more detail, we studied, first, the number of numerically significant elements ($\vartheta = 10^{-6}$) within the perturbed pseudo-density matrices. As shown in Table III the scaling behavior of the pseudo-densities perturbed with respect to \mathbf{m}_j is becoming sublinear in contrast to the linear-scaling behavior of the unperturbed pseudo-densities. Also, the transformations with these perturbed pseudo-densities become sublinear as shown in Table III. As a consequence, the number of transformed integrals, needed for the transformations with these perturbed pseudo-densities, become sublinear scaling as well. Similarly, the contractions to the perturbed R-matrices and those contractions involving the perturbed two-electron integrals scale sublinearly.

The scaling behavior of these steps is presented for calculating the shielding of the first and sixth C-atom of the alkane chain. As shown in Table III, the scaling behaviors are very similar, only the number of significant elements are higher for the central C-atom in the alkane chain.

Furthermore, we study the accuracy of ^{13}C -NMR shieldings as obtained by the numeric neglect of matrix elements (and integrals): For typical thresholds of $10^{-10}/10^{-8}/10^{-6}$, the differences to the exact value of the ^{13}C -NMR chemical shift for the first C-atom of $\text{C}_{20}\text{H}_{42}$ are $(-2.7 \times 10^{-5} \text{ ppm})/(-1.3 \times 10^{-2} \text{ ppm})/(2.7 \times 10^{-1} \text{ ppm})$ and for the sixth C-atom of $\text{C}_{20}\text{H}_{42}$ are $(-3.7 \times 10^{-5} \text{ ppm})/(-1.6 \times 10^{-3} \text{ ppm})/(-4.5 \times 10^{-3} \text{ ppm})$. The accuracy for the sixth C-atom is for the thresholds 10^{-8} and 10^{-6} higher due to the use of more integrals. Overall, this shows the possibility to fully control the numerical accuracy by changing the threshold.

2. Scaling behavior of the unperturbed R-matrices

By exploiting the local behavior of the nuclear magnetic spin perturbations, the formation of the unperturbed R-matrices can also be reduced to sublinear. While the sublinear steps discussed above can be straightforwardly screened for numerically significant contributions (as well as in the R-matrices perturbed with respect to \mathbf{m}_j), this is more difficult for truncating the unperturbed R-matrices. Although there are multiple pathways how to prescreen, we explore here a most simple first approach: In the derivation of the AO-MP2 shielding tensor presented above, the derivatives of the pseudo-densities are split into parts that depend on the different perturbed densities for applying the Z-vector method (Eq. (18)). However, if we write the chemical shielding tensor as the derivative of Eq. (A3) with respect to the nuclear magnetic moment:

$$\begin{aligned} \sigma_j^{MP2} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ 2 \left[\mathcal{I}^{\text{Bm}_j} + \sum_{\mu'\mu} \bar{R}_{\mu'\mu}^{\text{m}_j} \underline{P}_{\mu'\mu}^{\text{B}} \right. \right. \\ & + \sum_{\mu'\mu} \bar{R}_{\mu'\mu} \underline{P}_{\mu'\mu}^{\text{Bm}_j} + \sum_{v'v} \underline{R}_{v'v}^{\text{m}_j} \bar{P}_{v'v}^{\text{B}} \\ & \left. \left. + \sum_{v'v} \underline{R}_{v'v} \bar{P}_{v'v}^{\text{Bm}_j} \right] \right\}, \end{aligned} \quad (28)$$

the unperturbed matrices $\bar{\mathbf{R}}$ and \mathbf{R} are contracted with the second derivative of the pseudo-density matrix $\underline{P}^{\text{Bm}_j}$ and \bar{P}^{Bm_j} , respectively.

This has the advantage that the corresponding terms scale sublinearly as illustrated in Table V. In order to exploit this behavior we are deriving in the following a possible screening procedure for truncating the R-matrices.

We focus first on the term comprising the unperturbed matrix $\bar{\mathbf{R}}$. The occurring density derivative $\underline{P}^{\text{Bm}_j}$ can be written explicitly as

$$\begin{aligned} \underline{P}^{\text{Bm}_j} = & (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{m}_j} \mathbf{P}_{\text{occ}}^{\text{B}} + (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{B}} \mathbf{P}_{\text{occ}}^{\text{m}_j} + (\mathbf{e}^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{Bm}_j} \mathbf{P}_{\text{occ}} \\ & + (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}}) \mathbf{P}_{\text{occ}}^{\text{Bm}_j}. \end{aligned} \quad (29)$$

By applying cyclic permutations within the trace, we obtain

$$\begin{aligned} \text{Tr}[\bar{\mathbf{R}} \underline{P}^{\text{Bm}_j}] = & \text{Tr}[\bar{\mathbf{R}} (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{m}_j} \mathbf{P}_{\text{occ}}^{\text{B}}] + \text{Tr}[\mathbf{P}_{\text{occ}}^{\text{m}_j} \bar{\mathbf{R}} (\mathbf{e}^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{B}}] \\ & + \text{Tr}[\bar{\mathbf{R}} (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{Bm}_j} \mathbf{P}_{\text{occ}}] \\ & + \text{Tr}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j} \bar{\mathbf{R}} (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})], \end{aligned} \quad (30)$$

where the second derivative of the exponential term appears, which is

$$\begin{aligned} (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{Bm}_j} = & t_{\alpha} (\mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{F}^{\text{B}} + \mathbf{P}_{\text{occ}}^{\text{B}} \mathbf{F}^{\text{m}_j} + \mathbf{P}_{\text{occ}}^{\text{Bm}_j} \mathbf{F} + \mathbf{P}_{\text{occ}} \mathbf{F}^{\text{Bm}_j}) \\ & + \frac{1}{2!} t_{\alpha}^2 (\mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{F}^{\text{B}} \mathbf{P}_{\text{occ}} \mathbf{F} + \mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{F} \mathbf{P}_{\text{occ}}^{\text{B}} \mathbf{F} + \dots) \end{aligned} \quad (31)$$

In order to obtain a suitable approximation for screening the contributions to $\text{Tr}[\bar{\mathbf{R}} \underline{P}^{\text{Bm}_j}]$, we tested whether a restriction to the first term of the exponential expansion in Eq. (31) would be sufficient, since it is typically the largest term. Furthermore, the matrix \mathbf{F}^{Bm_j} is split into the part dependent on $\mathbf{P}_{\text{occ}}^{\text{Bm}_j}$ denoted by $\mathbf{G}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j}]$ and one independent on it, $\mathbf{F}^{(\text{Bm}_j)}$ (Eq. (13)). Inserting the truncated exponential expansion, one obtains

$$\begin{aligned} \text{Tr}[\bar{\mathbf{R}} (e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{Bm}_j} \mathbf{P}_{\text{occ}}] \\ \approx & \text{Tr} \left[t_{\alpha} (\bar{\mathbf{R}} \mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{F}^{\text{B}} \mathbf{P}_{\text{occ}} + \mathbf{F}^{\text{m}_j} \mathbf{P}_{\text{occ}} \bar{\mathbf{R}} \mathbf{P}_{\text{occ}}^{\text{B}} \right. \\ & + \bar{\mathbf{R}} \mathbf{P}_{\text{occ}}^{\text{Bm}_j} \mathbf{F} \mathbf{P}_{\text{occ}} + \bar{\mathbf{R}} \mathbf{P}_{\text{occ}} \mathbf{F}^{(\text{Bm}_j)} \mathbf{P}_{\text{occ}} \\ & \left. + \bar{\mathbf{R}} \mathbf{P}_{\text{occ}} \mathbf{G}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j}] \mathbf{P}_{\text{occ}} \right]. \end{aligned} \quad (32)$$

Similar to the derivation for the Z-vector equations for the second derivative of the density matrix $\mathbf{P}_{\text{occ}}^{\text{Bm}_j}$ described above, the term $\underline{\mathbf{A}}^{-1} \mathbf{b}^{\text{Bm}_j}$ can be inserted, and \mathbf{b}^{Bm_j} can be split into one part dependent on $\mathbf{P}_{\text{occ}}^{\text{B}}$ and in one independent on it: $\mathbf{b}^{\text{Bm}_j} = \mathbf{b}^{(\text{Bm}_j)} + \mathbf{b}[\mathbf{P}_{\text{occ}}^{\text{B}}]$. Thereby, the following expression

results:

$$\begin{aligned}
 \text{Tr}[\mathbf{R} \mathbf{P}^{\text{Bm}_j}] &\approx \underbrace{\text{Tr}[\mathbf{R} (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{m}_j} \mathbf{P}_{\text{occ}}^{\text{B}}]}_{(1)} + \underbrace{\text{Tr}[\mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{R} (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{B}}]}_{(2)} \\
 &+ \text{Tr} \left[t_\alpha \left(\underbrace{\mathbf{R} \mathbf{P}_{\text{occ}}^{\text{m}_j} \mathbf{F}^{\text{B}} \mathbf{P}_{\text{occ}}}_{(3)} + \underbrace{\mathbf{F}^{\text{m}_j} \mathbf{P}_{\text{occ}} \mathbf{R}}_{(4)} \mathbf{P}_{\text{occ}}^{\text{B}} \right. \right. \\
 &\quad \left. \left. + \underbrace{\mathbf{b}^{(\text{Bm}_j)} \mathbf{F} \mathbf{P}_{\text{occ}} \mathbf{R}}_{(5)} \mathbf{A}^{-1} + \underbrace{\mathbf{b}[\mathbf{P}_{\text{occ}}^{\text{B}}] \mathbf{F} \mathbf{P}_{\text{occ}} \mathbf{R}}_{(6)} \mathbf{A}^{-1} \right. \right. \\
 &\quad \left. \left. + \underbrace{\mathbf{R} \mathbf{P}_{\text{occ}} \mathbf{F}^{(\text{Bm}_j)} \mathbf{P}_{\text{occ}}}_{(7)} + \underbrace{\mathbf{R} \mathbf{P}_{\text{occ}} \mathbf{G}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j}] \mathbf{P}_{\text{occ}}}_{(8)} \right) \right] \\
 &+ \underbrace{\text{Tr}[\mathbf{A}^{-1} \mathbf{b}^{(\text{Bm}_j)} \mathbf{R} (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})]}_{(9)} + \underbrace{\text{Tr}[\mathbf{A}^{-1} \mathbf{b}[\mathbf{P}_{\text{occ}}^{\text{B}}] \mathbf{R} (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})]}_{(10)}.
 \end{aligned} \tag{33}$$

The same principle can be applied for the term $\text{Tr}[\mathbf{R} \mathbf{P}^{\text{Bm}_j}]$ by using \mathbf{P}_{virt} instead of \mathbf{P}_{occ} and the second derivative of the virtual density is $\mathbf{P}_{\text{virt}}^{\text{Bm}_j} = -\mathbf{P}_{\text{occ}}^{\text{Bm}_j}$ (derivative of Eq. (B4) in Appendix B). With this the terms for the virtual part result as

$$\begin{aligned}
 \text{Tr}[\mathbf{R} \mathbf{P}^{\text{Bm}_j}] &\approx \underbrace{\text{Tr}[\mathbf{R} (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^{\text{m}_j} \mathbf{P}_{\text{virt}}^{\text{B}}]}_{(1)} + \underbrace{\text{Tr}[\mathbf{P}_{\text{virt}}^{\text{m}_j} \mathbf{R} (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^{\text{B}}]}_{(2)} \\
 &+ \text{Tr} \left[t_\alpha \left(\underbrace{\mathbf{R} \mathbf{P}_{\text{virt}}^{\text{m}_j} \mathbf{F}^{\text{B}} \mathbf{P}_{\text{virt}}}_{(3)} + \underbrace{\mathbf{F}^{\text{m}_j} \mathbf{P}_{\text{virt}} \mathbf{R}}_{(4)} \mathbf{P}_{\text{virt}}^{\text{B}} \right. \right. \\
 &\quad \left. \left. - \underbrace{(\mathbf{b}^{(\text{Bm}_j)} \mathbf{F} \mathbf{P}_{\text{virt}} \mathbf{R}) \mathbf{A}^{-1}}_{(5)} - \underbrace{\mathbf{b}[\mathbf{P}_{\text{occ}}^{\text{B}}] \mathbf{F} \mathbf{P}_{\text{virt}} \mathbf{R} \mathbf{A}^{-1}}_{(6)} \right. \right. \\
 &\quad \left. \left. + \underbrace{\mathbf{R} \mathbf{P}_{\text{virt}} \mathbf{F}^{(\text{Bm}_j)} \mathbf{P}_{\text{virt}}}_{(7)} + \underbrace{\mathbf{R} \mathbf{P}_{\text{virt}} \mathbf{G}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j}] \mathbf{P}_{\text{virt}}}_{(8)} \right) \right] \\
 &- \underbrace{\text{Tr}[\mathbf{A}^{-1} \mathbf{b}^{(\text{Bm}_j)} \mathbf{R} (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})]}_{(9)} - \underbrace{\text{Tr}[\mathbf{A}^{-1} \mathbf{b}[\mathbf{P}_{\text{occ}}^{\text{B}}] \mathbf{R} (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})]}_{(10)}.
 \end{aligned} \tag{34}$$

For preselecting significant elements of the R-matrices contributing to the final trace expressions on the left-hand side of Eqs. (33) and (34), we have selected ten matrix products marked with a brace, that could be used for screening

TABLE IV. Screening influence on ^{13}C -NMR shieldings for the first C-atom of alkanes (STO-3G basis) as compared to the exact value (deviation in ppm) by using five terms of the Eqs. (33) and (34) (labeled 5 \times) and ten terms (denoted 10 \times). For a description of the approximations see text. The first threshold applies to selecting significant elements in the matrix product and the second to used integral products $[(\mu' \bar{\nu} | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)]$ and $[(\mu \nu' | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)]$. The data apply to one Laplace point.

	$\text{C}_{15}\text{H}_{32}$		$\text{C}_{20}\text{H}_{42}$	
	5 \times	10 \times	5 \times	10 \times
$10^{-8}/10^{-8}$	1.2×10^{-4}	1.2×10^{-4}	-2.0×10^{-4}	-2.0×10^{-4}
$10^{-6}/10^{-6}$	1.7×10^{-2}	1.7×10^{-2}	1.6×10^{-2}	1.6×10^{-2}
$10^{-4}/10^{-4}$	-1.0×10^0	-1.0×10^0	-9.7×10^{-1}	-9.7×10^{-1}
$10^{-4}/10^{-6}$	3.1×10^{-2}	3.2×10^{-2}	3.3×10^{-2}	3.4×10^{-2}
$10^{-2}/10^{-6}$	-6.0×10^{-1}	3.8×10^{-1}	-9.5×10^{-1}	1.5×10^{-1}

as investigated below. Once the contributing R-elements have been identified, they can be build via the corresponding integral products $\bar{R}_{\mu'\mu} = \sum_{\nu\lambda\sigma} (\mu' \bar{\nu} | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)$ and $R_{\nu'\nu} = \sum_{\mu\lambda\sigma} (\mu \nu' | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)$. Here, the numerically significant integral products can be preselected in the future by, e.g., using the recently introduced QQR estimates.^{37,38}

In order to study the accuracy of employing the ten matrix products (denoted as 10 \times screening) described above for screening the numerically significant elements in $\bar{\mathbf{R}}$ and \mathbf{R} , we list in Table IV the influence on the final NMR shielding tensor for various thresholds. Here, in addition, the integral products $[(\mu' \bar{\nu} | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)]$ and $[(\mu \nu' | \underline{\lambda} \bar{\sigma})(\mu \nu | \lambda \sigma)]$ have been screened with the same threshold. The data show the systematic increase of the accuracy in the NMR shieldings with tighter thresholds and that quickly accuracies of better than 0.1 ppm are reached, although much more extensive studies are required in future work.

While the ten-terms estimate seems to work sufficiently well, such a screening would be rather inefficient, since, in particular, matrices like $\mathbf{P}_{\text{occ}}^{\text{B}}$ and $\mathbf{P}_{\text{occ}}^{\text{Bm}_j}$ are normally not calculated and avoided using the Z-vector method. Therefore, we have reduced the number of terms to include just five

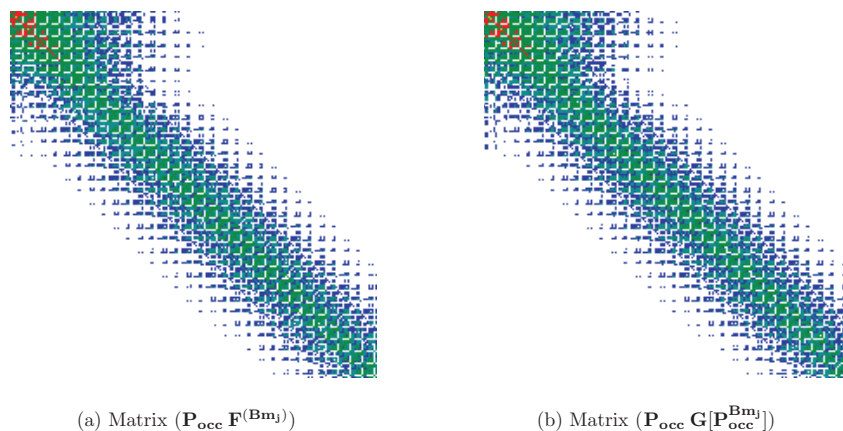


FIG. 2. Similarities of the sparsity patterns of the matrices $(\mathbf{P}_{\text{occ}} \mathbf{F}^{(\text{Bm}_j)})$ and $(\mathbf{P}_{\text{occ}} \mathbf{G}[\mathbf{P}_{\text{occ}}^{\text{Bm}_j}])$ for the molecule $\text{C}_{30}\text{H}_{62}$ (STO-3G basis). Elements are signed by different colours dependent on their sizes: value $>10^0$ yellow, $>10^{-1}$ red, $>10^{-3}$ green, $>10^{-4}$ cyan, $>10^{-5}$ blue, and $>10^{-6}$ white.

matrix multiplications (marked by a circle in Eqs. (33) and (34)) for preselecting the significant elements of $\bar{\mathbf{R}}$ and \mathbf{R} . The accuracy of such a five-terms screening (5 \times -screening) is only slightly reduced as compared to the ten-terms screening as shown in Table IV and seems fully controlled by the numerical threshold.

To study the difference between the 5 \times - and 10 \times -screening procedure in more detail, we calculated the different number of significant elements in the terms forming the R-matrices using a threshold of $\vartheta = 10^{-6}$. For example, for the $\text{C}_{20}\text{H}_{42}$ molecule and the R-matrix \mathbf{R} , the influence on matrix products *not* included in the screening process is rather small: 1.3% of the elements in matrix product $\mathbb{8}$ and 0.2 % in matrix product $\mathbb{10}$ are lost by using the 5 \times -screening, while for the other three matrix products no elements are lost in the R-matrix. At the same time, it has to be stressed that the largest matrix product neglected is 1.6×10^{-6} , which is only slightly larger than the chosen threshold. This reflects the fact that the matrices influenced by the same perturbation have a similar decay behavior as illustrated by the similar patterns shown in Figure 2. This similar decay is the reason, why it seems not necessary to employ all ten matrices for screening.

The 5 \times -screening represents a first step towards a possible screening of contributions, so that the scaling behavior of the remaining significant elements in the R-matrices and in the integral contraction steps in forming them is becoming sublinear as shown in Table V. For the presented small sys-

TABLE V. The first two rows show the number of significant elements and scaling behavior (in parentheses) of the contributing products within the matrix multiplication of the second derivative of the pseudo-densities with the R-matrices (see Eq. (28)). Furthermore, the effect of a preliminary preselection using the 5 \times -screening (discussed in text) is shown: The number of preselected significant elements and scaling behavior (in parentheses) in the R-matrices and in the contraction products of two-electron integrals to form the R-matrices [Eq. (7)] is listed for different thresholds (^{13}C -NMR shieldings of the first C-atom of different alkanes; one Laplace point; basis set STO-3G).

	$\text{C}_{15}\text{H}_{32}$	$\text{C}_{20}\text{H}_{42}$		$\text{C}_{30}\text{H}_{62}$	
$\sum_i \bar{\mathbf{R}}_{\mu i} \mathbf{P}_{i\mu}^{\text{Bm}} (10^{-6})$	48 039	53 194	(0.4)	61 210	(0.4)
$\sum_i \mathbf{R}_{\nu i} \bar{\mathbf{P}}_{i\nu}^{\text{Bm}} (10^{-6})$	85 242	96 290	(0.4)	114 677	(0.4)
$\bar{\mathbf{R}}$ (matrix)					
10^{-2}	105	112	(0.2)	118	(0.1)
10^{-4}	1643	2042	(0.8)	2745	(0.7)
10^{-6}	3925	5197	(1.0)	7666	(1.0)
$\bar{\mathbf{R}}$ (contraction)					
10^{-2}	1576	1756	(0.4)	1913	(0.2)
10^{-4}	745 897	966 285	(0.9)	1 370 735	(0.9)
10^{-6}	19 384 967	26 801 708	(1.1)	41 400 900	(1.1)
\mathbf{R} (matrix)					
10^{-2}	250	265	(0.2)	293	(0.3)
10^{-4}	2399	3162	(1.0)	4641	(1.0)
10^{-6}	4130	5501	(1.0)	8158	(1.0)
\mathbf{R} (contraction)					
10^{-2}	1006	1133	(0.4)	1353	(0.4)
10^{-4}	769 782	1 033 696	(1.0)	1 552 135	(1.0)
10^{-6}	18 777 858	26 035 351	(1.2)	40 381 290	(1.1)

TABLE VI. Number of significant elements and scaling behavior (in parentheses) of the Z-vector for avoiding \mathbf{P}^{B} and the contractions with the two-electron integrals (\mathbf{G}). For comparison also the unfavorable scaling behavior of \mathbf{P}^{B} and of its contractions is shown, which are avoided in our present formulation (^{13}C -NMR shieldings of the first C-atom of different alkanes; one Laplace point; $\vartheta = 10^{-6}$; basis set STO-3G; perturbation $\mathbf{m}_{j,x}$ and \mathbf{B}_x).

	$\text{C}_{15}\text{H}_{32}$	$\text{C}_{20}\text{H}_{42}$	
$\mathbf{Z}_{\text{ov}}^{\text{mj}} + \mathbf{Z}_{\text{vo}}^{\text{mj}}$	2232	2315	(0.1)
$\mathbf{G}[\mathbf{Z}_{\text{ov}}^{\text{mj}} + \mathbf{Z}_{\text{vo}}^{\text{mj}}]$	207 391	234 142	(0.4)
\mathbf{P}^{B}	7326	12 418	(1.9)
$\mathbf{G}[\mathbf{P}_{\text{ov}}^{\text{B}} + \mathbf{P}_{\text{vo}}^{\text{B}}]$	7 645 442	13 407 384	(2.0)

tems the scaling behavior is linear for tighter thresholds, but is becoming sublinear for less tight ones. This is a strong hint for the scaling behavior of larger molecules. For comparison, if one considers, e.g., the density matrix and would loosen the screening thresholds, the scaling would be at best linear: For the molecules $\text{C}_{20}\text{H}_{42}$ and $\text{C}_{25}\text{H}_{52}$ (basis set STO-3G), e.g., the scaling behavior of the density matrix for different thresholds $10^{-6}/10^{-4}/10^{-2}$ is $(1.4)/(1.2)/(1.1)$.

In contrast, if matrices are locally perturbed (by the nuclear-magnetic moment), then sublinear scaling occurs. While for small molecules this is only visible for loose thresholds, we expect this to hold also for tighter thresholds if larger molecules are considered.

While the present screening procedure is only a first idea for preselecting significant contributions, much more extensive testing is required in the future. Nevertheless, the first results seem promising for developing an efficient screening with controlled accuracy.

3. Scaling behavior of the Z-vector equations for the first and second derivative

The most expensive step in the AO-based Z-vector method is the contraction of the Z-vector with the two-electron integrals, which scales formally with N^4 , but can be reduced to linear or even sublinear by exploiting locality. The data shown in Table VI illustrate that the Z-vector for the first derivative and the contraction with the two-electron integrals

TABLE VII. Number of significant elements and scaling behavior (in parentheses) of the Z-vector employed for avoiding \mathbf{P}^{Bm} and the contractions with the two-electron integrals. Different thresholds are listed (^{13}C -NMR shieldings of the first C-atom of different alkanes, one Laplace point; basis set STO-3G; thresholds for the R-matrices: 10^{-2} for the matrices, 10^{-6} for the contractions).

	$\text{C}_{20}\text{H}_{42}$	$\text{C}_{25}\text{H}_{52}$		$\text{C}_{30}\text{H}_{62}$	
$\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}$					
10^{-2}	170	196	(0.6)	216	(0.5)
10^{-4}	6448	8140	(1.1)	9622	(0.9)
10^{-6}	12 716	17 207	(1.4)	21 712	(1.3)
$\mathbf{G}[\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}]$					
10^{-2}	2206	2580	(0.7)	2720	(0.3)
10^{-4}	408 477	542 254	(1.3)	649 364	(1.0)
10^{-6}	5 735 796	8 124 522	(1.6)	10 346 937	(1.3)

can also become sublinear. This shows again the important advantage of our present formulation as the Z-vector method allows to avoid the formation of the linear-scaling quantity $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$. A similar sublinear scaling holds for the Z-vector equations for the second derivative by using the earlier introduced truncation of the R-matrices (see Table VII). For these small molecules, the sublinear scaling occurs only for less tight thresholds, but as explained in Sec. III C 2, this is a strong indication for sublinear scaling for larger molecules.

IV. CONCLUSION

We have presented an AO-based reformulation for calculating NMR chemical shielding tensors at the MP2 level that reduces the asymptotic scaling behavior from fifth power to linear or even sublinear for selected nuclei in the rate-determining steps. Here, we employ the Laplace formulation for avoiding MO quantities as a starting point, so that only local quantities occur and numerically insignificant elements can be screened and avoided, e.g., by our recently introduced QQR integral estimates.^{37,38} The key for achieving sublinear scaling is to avoid global perturbations with respect to the magnetic field and, instead, to solve for perturbations with respect to the nuclear magnetic moment. A suitable formulation is obtained by employing an AO-based Z-vector formulation. In this way, the chemical shielding tensor of a specific nucleus can be obtained in an asymptotically sublinear-scaling way in the rate-determining steps. A similar ansatz was applied for avoiding the second derivative of the density matrix. The corresponding AO-based Z-vector method within a density matrix-based scheme presented in this work scales also sublinearly by using a simple approximation for truncating the R-matrices.

While our pilot implementation is still constrained to small molecules and basis sets, our first results show the validity of all equations. Furthermore, the required number of Laplace points for reliable NMR shieldings lies typically between three and four, which is in line with the observations made for the AO-MP2 energies and gradients, while NMR shieldings seem even faster converging. In addition, we studied the scaling of the transformation and contraction steps and confirm the expected sublinear behavior, while the accuracy of the results can be fully controlled by the screening threshold.

Overall, we expect our method to be highly useful for the reliable calculation of NMR chemical shieldings at the MP2 level not only for large molecular systems, but also in describing influences by an explicit solvent environment, where, in particular, the chemical shieldings for just a few nuclei are of importance.

ACKNOWLEDGMENTS

The authors thank Dr. M. Beer, Dr. J. Kussmann, and S. A. Maurer (University of Munich, LMU) for useful discussions. C.O. acknowledges financial support by the Volkswagen Stiftung within the funding initiative “New Conceptual Approaches to Modeling and Simulation of Complex Systems,” by the SFB 749 “Dynamik und Intermediate moleku-

larer Transformationen” (DFG), and the DFG cluster of excellence (EXC 114) “Center for Integrative Protein Science Munich” (CIPSM).

APPENDIX A: THE FIRST DERIVATIVE WITH RESPECT TO THE MAGNETIC FIELD B

Similar to the AO-MP2 gradients, introduced by Schweizer *et al.* in 2009,⁴¹ the first derivative of the energy with respect to the magnetic field B is given by

$$E_{\text{AO-MP2}}^{\mathbf{B}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ 2 \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}} + 2 \sum_{\mu\nu\lambda\sigma} \sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{B}} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma) + 2 \sum_{\mu\nu\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{B}} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma) \right\}. \quad (\text{A1})$$

By defining the following matrices:

$$\begin{aligned} \mathcal{I}^{\mathbf{B}} &= \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu||\lambda\sigma)^{\mathbf{B}}, \\ \bar{R}_{\mu'\mu} &= \sum_{\nu\lambda\sigma} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})[2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)], \\ \underline{R}_{\nu'\nu} &= \sum_{\mu\lambda\sigma} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})[2(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)], \end{aligned} \quad (\text{A2})$$

the expression for the derivative of the energy simplifies to

$$E_{\text{AO-MP2}}^{\mathbf{B}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ 2 \left[\mathcal{I}^{\mathbf{B}} + \sum_{\mu'\mu} \bar{R}_{\mu'\mu} \underline{P}_{\mu'\mu}^{\mathbf{B}} + \sum_{\nu'\nu} \underline{R}_{\nu'\nu} \bar{P}_{\nu'\nu}^{\mathbf{B}} \right] \right\}. \quad (\text{A3})$$

The first term includes only derivatives of the two electron integrals, which can be cheaply calculated. In contrast, the explicit calculation of the derivative of the pseudo-densities is expensive and can be avoided as shown in Appendix A 1.

1. Avoiding the computation of $\underline{P}^{\mathbf{B}}$ and $\bar{P}^{\mathbf{B}}$

In Eq. (3) the pseudo-densities are defined using the coefficient matrices. While the SCF diagonalization overhead for generating the coefficient matrices is small, for a fully linear scaling method the coefficient matrices need to be avoided. Therefore, the pseudo-densities are formulated in terms of the sparse density matrix instead.⁴²

$$\underline{P}_{\mu'\mu} = \sum_i C_{\mu'i} e^{\varepsilon_i t_a} C_{\mu i} = (e^{t_a \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}})_{\mu'\mu}, \quad (\text{A4a})$$

$$\bar{P}_{\nu'\nu} = \sum_a C_{\nu'a} e^{-\varepsilon_a t_a} C_{\nu a} = (e^{-t_a \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}})_{\nu'\nu}. \quad (\text{A4b})$$

Here, \mathbf{P}_{occ} is the occupied, \mathbf{P}_{virt} the virtual density matrix, and \mathbf{F} the Fock matrix.

The derivatives of the pseudo-densities are given by

$$\begin{aligned}\underline{P}_{\mu'\mu}^{\mathbf{B}} &= (e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{occ}} + e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}}^{\mathbf{B}}, \\ \underline{P}_{\nu'\nu}^{\mathbf{B}} &= (e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{virt}} + e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}}^{\mathbf{B}}.\end{aligned}\quad (\text{A5})$$

The exponentials $e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}}$ and $e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}}$ used in this equation are matrix exponentials and can be obtained by a Taylor series:

$$e^{\mathbf{A}} = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{A}^k = 1 + \mathbf{A} + \frac{1}{2!} \mathbf{A}^2 + \dots \quad (\text{A6})$$

The same principle can be applied to the derivative of the exponentials:

$$\begin{aligned}(e^{\mathbf{A}})^{\mathbf{B}} &= \sum_{k=0}^{\infty} \frac{1}{k!} (\mathbf{A}^k)^{\mathbf{B}} \\ &= 0 + \mathbf{A}^{\mathbf{B}} + \frac{1}{2!} (\mathbf{A}^{\mathbf{B}} \mathbf{A} + \mathbf{A} \mathbf{A}^{\mathbf{B}}) \\ &\quad + \frac{1}{3!} (\mathbf{A}^{\mathbf{B}} \mathbf{A}^2 + \mathbf{A} \mathbf{A}^{\mathbf{B}} \mathbf{A} + \mathbf{A}^2 \mathbf{A}^{\mathbf{B}}) + \dots\end{aligned}\quad (\text{A7})$$

In the following, the Taylor series will be used to calculate the derivative of the pseudo-densities efficiently. To accomplish this, the first step is to insert the definition with the exponentials from Eq. (A5) in Eq. (A3):

$$\begin{aligned}E_{\text{AO-MP2}}^{\mathbf{B}} &= - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ 2Z^{\mathbf{B}} + 2\text{Tr} \left[\underbrace{\underline{\mathbf{R}}((e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{occ}} + e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}} \mathbf{P}_{\text{occ}}^{\mathbf{B}})}_{\mathcal{A}_1} \right] \right. \\ &\quad \left. + 2\text{Tr} \left[\underbrace{\underline{\mathbf{R}}((e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{virt}} + e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}} \mathbf{P}_{\text{virt}}^{\mathbf{B}})}_{\mathcal{B}_2} \right] \right\}.\end{aligned}\quad (\text{A8})$$

Here, the derivative $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ appears, which is normally expensive to calculate. However, by using the Z-vector-method of Handy and Schaefer,^{39,40} it can be avoided. Furthermore, the derivative $\mathbf{P}_{\text{virt}}^{\mathbf{B}}$ can be substituted by an expression, that depends on $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$, as will be explained later.

To exploit the Z-vector-method, the expression must have the following form:

$$E_{\text{AO-MP2}}^{\mathbf{B}} = \text{Tr}[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{B}}] + \text{Tr}[\mathcal{X}], \quad (\text{A9})$$

where the first term is dependent on $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ and the second one is independent.

The terms \mathcal{B}_1 and \mathcal{B}_2 in Eq. (A8) are already in the correct form, in contrast to the terms \mathcal{A}_1 and \mathcal{A}_2 (Eq. (A8)). The latter can be rearranged by inserting the derivative of the expansions of the exponentials of Eq. (A7) and by applying cyclic permutations within the trace to move $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ and $\mathbf{P}_{\text{virt}}^{\mathbf{B}}$ to the right-hand side within the trace.

Furthermore, the derivative of the Fock matrix $\mathbf{F}^{\mathbf{B}}$ will be split into a part dependent on $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ and an independent part labeled $\mathbf{F}^{(\mathbf{B})}$:

$$\mathbf{F}^{\mathbf{B}} = \underbrace{\mathbf{H}^{\mathbf{B}} + \mathbf{G}^{\mathbf{B}}[\mathbf{P}_{\text{occ}}]}_{\mathbf{F}^{(\mathbf{B})}} + \mathbf{G}[\mathbf{P}_{\text{occ}}^{\mathbf{B}}]. \quad (\text{A10})$$

Here, $\mathbf{G}[\mathbf{P}_{\text{occ}}^{\mathbf{B}}]$ is the contraction of the two-electron integrals with the perturbed density matrix and $\mathbf{G}^{\mathbf{B}}[\mathbf{P}_{\text{occ}}]$ is the

contraction of the derivative of the two-electron integrals with the unperturbed density matrix. After insertion into term \mathcal{A}_1 , one obtains

$$\text{Tr}[\underline{\mathbf{R}}(e^{t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{occ}}] = \text{Tr}[(\bar{\mathbf{Y}}_1 + \mathbf{G}[\bar{\mathbf{Y}}_2]) \mathbf{P}_{\text{occ}}^{\mathbf{B}}] + \text{Tr}[\bar{\mathbf{Y}}_2 \mathbf{F}^{(\mathbf{B})}]. \quad (\text{A11})$$

$\bar{\mathbf{Y}}_1 = \sum_{k=1}^m \bar{\mathbf{Y}}_1^{(k)}$ and $\bar{\mathbf{Y}}_2 = \sum_{k=1}^m \bar{\mathbf{Y}}_2^{(k)}$ are recursion formulae, where each term can be built from

$$\bar{\mathbf{Y}}_1^{(k)} = \frac{1}{k} [(t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}} \bar{\mathbf{R}})(t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(k-1)} + (t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}}) \bar{\mathbf{Y}}_1^{(k-1)}], \quad (\text{A12a})$$

$$\bar{\mathbf{Y}}_2^{(k)} = \frac{1}{k} [(t_\alpha \mathbf{P}_{\text{occ}} \bar{\mathbf{R}} \mathbf{P}_{\text{occ}})(t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}})^{(k-1)} + (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}) \bar{\mathbf{Y}}_2^{(k-1)}], \quad (\text{A12b})$$

while it starts with $\bar{\mathbf{Y}}_1^{(0)} = 0$ and $\bar{\mathbf{Y}}_2^{(0)} = 0$.

Similarly, one obtains the terms for \mathcal{A}_2 (Eq. (A8))

$$\begin{aligned}\text{Tr}[\underline{\mathbf{R}}(e^{-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}})^{\mathbf{B}} \mathbf{P}_{\text{virt}}] &= \text{Tr}[\underline{\mathbf{Y}}_1 \mathbf{P}_{\text{virt}}^{\mathbf{B}}] + \text{Tr}[\underline{\mathbf{Y}}_2 \mathbf{F}^{\mathbf{B}}] \\ &\quad + \text{Tr}[\mathbf{G}[\underline{\mathbf{Y}}_2] \mathbf{P}_{\text{occ}}^{\mathbf{B}}],\end{aligned}\quad (\text{A13})$$

where the $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ -dependent term is part of the derivative of the Fock matrix $\mathbf{F}^{\mathbf{B}}$.

$\underline{\mathbf{Y}}_1 = \sum_{k=1}^m \underline{\mathbf{Y}}_1^{(k)}$ and $\underline{\mathbf{Y}}_2 = \sum_{k=1}^m \underline{\mathbf{Y}}_2^{(k)}$ are again recursion formulae:

$$\underline{\mathbf{Y}}_1^{(k)} = \frac{1}{k} [(-t_\alpha \mathbf{F} \mathbf{P}_{\text{virt}} \underline{\mathbf{R}})(-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F})^{(k-1)} + (-t_\alpha \mathbf{F} \mathbf{P}_{\text{virt}}) \underline{\mathbf{Y}}_1^{(k-1)}], \quad (\text{A14a})$$

$$\begin{aligned}\underline{\mathbf{Y}}_2^{(k)} &= \frac{1}{k} [(-t_\alpha \mathbf{P}_{\text{virt}} \underline{\mathbf{R}} \mathbf{P}_{\text{virt}})(-t_\alpha \mathbf{F} \mathbf{P}_{\text{virt}})^{(k-1)} \\ &\quad + (-t_\alpha \mathbf{P}_{\text{virt}} \mathbf{F}) \underline{\mathbf{Y}}_2^{(k-1)}],\end{aligned}\quad (\text{A14b})$$

and the expression also begins with $\underline{\mathbf{Y}}_1^{(0)} = 0$ and $\underline{\mathbf{Y}}_2^{(0)} = 0$.

As mentioned earlier, the perturbed virtual density matrix can be substituted by an expression dependent on the perturbed occupied density matrix. The following relation between them holds for a non-orthogonal basis:

$$\mathbf{P}_{\text{occ}} \mathbf{S} + \mathbf{P}_{\text{virt}} \mathbf{S} = \mathbf{1}. \quad (\text{A15})$$

By differentiating this expression:

$$\mathbf{P}_{\text{occ}}^{\mathbf{B}} \mathbf{S} + \mathbf{P}_{\text{occ}} \mathbf{S}^{\mathbf{B}} + \mathbf{P}_{\text{virt}}^{\mathbf{B}} \mathbf{S} + \mathbf{P}_{\text{virt}} \mathbf{S}^{\mathbf{B}} = \mathbf{0}, \quad (\text{A16})$$

multiplying from the right with \mathbf{S}^{-1} :

$$\mathbf{P}_{\text{occ}}^{\mathbf{B}} + \mathbf{P}_{\text{occ}} \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1} + \mathbf{P}_{\text{virt}}^{\mathbf{B}} + \mathbf{P}_{\text{virt}} \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1} = \mathbf{0}, \quad (\text{A17})$$

and by solving the equation for $\mathbf{P}_{\text{virt}}^{\mathbf{B}}$:

$$\mathbf{P}_{\text{virt}}^{\mathbf{B}} = -\mathbf{P}_{\text{occ}}^{\mathbf{B}} - \underbrace{(\mathbf{P}_{\text{occ}} + \mathbf{P}_{\text{virt}}) \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1}}_{\mathbf{S}^{-1}}, \quad (\text{A18})$$

one can replace the derivative of the unoccupied density by this expression.

Finally, the expression for the first derivative with respect to the magnetic field \mathbf{B} is

$$\begin{aligned}
 E_{\text{AO-MP2}}^{\mathbf{B}} &= - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2\mathcal{I}^{\mathbf{B}} \\
 &\quad + 2\text{Tr}[(\bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G}[\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] \\
 &\quad + \bar{\mathbf{R}}e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} - \underline{\mathbf{R}}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})\mathbf{P}_{\text{occ}}^{\mathbf{B}}] \\
 &\quad + 2\text{Tr}[(\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2)\mathbf{F}^{(\mathbf{B})}] \\
 &\quad + 2\text{Tr}[-(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1}] \} \\
 &= - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2\mathcal{I}^{\mathbf{B}} + 2\text{Tr}[\mathcal{P}\mathbf{P}_{\text{occ}}^{\mathbf{B}}] \\
 &\quad + 2\text{Tr}[\mathcal{F}\mathbf{F}^{(\mathbf{B})}] + 2\text{Tr}[\mathcal{S}\mathbf{S}^{-1}\mathbf{S}^{\mathbf{B}}\mathbf{S}^{-1}] \}, \quad (\text{A19})
 \end{aligned}$$

with

$$\mathcal{P} = \bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G}[\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] + \bar{\mathbf{R}}e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} - \underline{\mathbf{R}}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}}, \quad (\text{A20a})$$

$$\mathcal{F} = \bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2, \quad (\text{A20b})$$

$$\mathcal{S} = -(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}}e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}}). \quad (\text{A20c})$$

APPENDIX B: DERIVATIVE OF THE RECURSION FORMULAE $\bar{\mathbf{Y}}_1$, $\underline{\mathbf{Y}}_1$, $\bar{\mathbf{Y}}_2$, AND $\underline{\mathbf{Y}}_2$

The terms in Eqs. (11) and (14) contain the derivatives of the recursion formulae $\bar{\mathbf{Y}}_1$ and $\underline{\mathbf{Y}}_1$. To obtain them efficiently, one has to reformulate the derivative in a recursive form again. By differentiating and separating each term that includes either \mathbf{P}^{mj} , \mathbf{F}^{mj} , or \mathbf{R}^{mj} , one obtains the following new recursion formulae for the derivatives:

$$\begin{aligned}
 \bar{\mathbf{Y}}_1^{\text{mj}} &= \bar{\mathbf{D}} + \bar{\mathbf{E}} + \bar{\mathbf{J}}, \\
 \underline{\mathbf{Y}}_1^{\text{mj}} &= \underline{\mathbf{D}} + \underline{\mathbf{E}} + \underline{\mathbf{J}}.
 \end{aligned} \quad (\text{B1})$$

Each term of the new recursion formulae $\bar{\mathbf{D}} = \sum_{k=1}^m \bar{\mathbf{D}}^{(k)}$, $\bar{\mathbf{E}} = \sum_{k=1}^m \bar{\mathbf{E}}^{(k)}$, and $\bar{\mathbf{J}} = \sum_{k=1}^m \bar{\mathbf{J}}^{(k)}$ is calculated by

$$\begin{aligned}
 \bar{\mathbf{D}}^{(k)} &= \frac{1}{k} \left[\bar{\mathbf{D}}^{(k-1)} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}) \right. \\
 &\quad + \sum_{i=0}^{k-1} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(i)} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}}^{\text{mj}}) (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(k-1-i)} \bar{\mathbf{R}} \\
 &\quad \left. + \sum_{i=0}^{k-2} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(k-1-i)} \bar{\mathbf{R}} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(i)} (t_{\alpha}\mathbf{P}_{\text{occ}}^{\text{mj}}\mathbf{F}) \right], \quad (\text{B2a})
 \end{aligned}$$

$$\begin{aligned}
 \bar{\mathbf{E}}^{(k)} &= \frac{1}{k} \left[\bar{\mathbf{E}}^{(k-1)} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}) \right. \\
 &\quad + \sum_{i=0}^{k-1} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(i)} \mathbf{F}^{\text{mj}} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(k-1-i)} (t_{\alpha}\mathbf{P}_{\text{occ}} \bar{\mathbf{R}}) \\
 &\quad \left. + \sum_{i=0}^{k-2} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(k-1-i)} \bar{\mathbf{R}} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(i)} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}^{\text{mj}}) \right], \quad (\text{B2b})
 \end{aligned}$$

$$\bar{\mathbf{J}}^{(k)} = \frac{1}{k} [\bar{\mathbf{J}}^{(k-1)} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}) + (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(k)} \bar{\mathbf{R}}^{\text{mj}}], \quad (\text{B2c})$$

while the recursions start with $\bar{\mathbf{D}}^{(0)} = \mathbf{0}$, $\bar{\mathbf{E}}^{(0)} = \mathbf{0}$, and $\bar{\mathbf{J}}^{(0)} = \mathbf{0}$.

The virtual parts $\underline{\mathbf{D}}$, $\underline{\mathbf{E}}$, and $\underline{\mathbf{J}}$ can be built in the same way, with the only difference of using \mathbf{P}_{virt} instead of \mathbf{P}_{occ} and $\underline{\mathbf{R}}$ instead of $\bar{\mathbf{R}}$. Therefore, the derivative of \mathbf{P}_{virt} with respect to \mathbf{m}_j appears, which can, similar to the derivative with respect to the magnetic field, be substituted by an expression containing the perturbed occupied density matrix only: The derivative of Eq. (A15) is

$$\mathbf{P}_{\text{occ}}^{\text{mj}} \mathbf{S} + \mathbf{P}_{\text{virt}}^{\text{mj}} \mathbf{S} = \mathbf{0}. \quad (\text{B3})$$

Here, in contrast to the derivative of the overlap matrix with respect to the magnetic field, the derivative with respect to the nuclear magnetic moment is zero: $\mathbf{S}^{\text{mj}} = \mathbf{0}$. By multiplying Eq. (B3) from the right with \mathbf{S}^{-1} , the following relation between the perturbed occupied and unoccupied density matrices results:

$$\mathbf{P}_{\text{virt}}^{\text{mj}} = -\mathbf{P}_{\text{occ}}^{\text{mj}}. \quad (\text{B4})$$

For the terms (11) and (12) the derivatives of $\bar{\mathbf{Y}}_2$ and $\underline{\mathbf{Y}}_2$ are necessary. Similar to the terms described above, they can be obtained by new recursion formulae:

$$\begin{aligned}
 \bar{\mathbf{Y}}_2^{\text{mj}} &= \bar{\mathbf{M}} + \bar{\mathbf{K}} + \bar{\mathbf{L}}, \\
 \underline{\mathbf{Y}}_2^{\text{mj}} &= \underline{\mathbf{M}} + \underline{\mathbf{K}} + \underline{\mathbf{L}},
 \end{aligned} \quad (\text{B5})$$

with $\bar{\mathbf{M}} = \sum_{k=1}^m \bar{\mathbf{M}}^{(k)}$, $\bar{\mathbf{K}} = \sum_{k=1}^m \bar{\mathbf{K}}^{(k)}$, $\bar{\mathbf{L}} = \sum_{k=1}^m \bar{\mathbf{L}}^{(k)}$, and the recursions terms given by

$$\begin{aligned}
 \bar{\mathbf{M}}^{(k)} &= \frac{1}{k} \left[\bar{\mathbf{M}}^{(k-1)} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}}) \right. \\
 &\quad + \sum_{i=0}^{k-1} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(i)} \mathbf{P}_{\text{occ}}^{\text{mj}} (t_{\alpha}\mathbf{F}\mathbf{P}_{\text{occ}})^{(k-1-i)} (t_{\alpha}\bar{\mathbf{R}}\mathbf{P}_{\text{occ}}) \\
 &\quad \left. + \sum_{i=0}^{k-1} (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(i)} (t_{\alpha}\mathbf{P}_{\text{occ}}\bar{\mathbf{R}}) (t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F})^{(k-1-i)} \mathbf{P}_{\text{occ}}^{\text{mj}} \right], \quad (\text{B6a})
 \end{aligned}$$

$$\begin{aligned} \bar{\mathbf{K}}^{(k)} = & \frac{1}{k} \left[\bar{\mathbf{K}}^{(k-1)} (t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}}) \right. \\ & + \sum_{i=0}^{k-2} (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(i)} \mathbf{P}_{\text{occ}} \bar{\mathbf{R}} (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(k-2-i)} (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F}^{\text{mj}} \mathbf{P}_{\text{occ}}) \\ & \left. + \sum_{i=0}^{k-2} (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(i)} \mathbf{P}_{\text{occ}} \mathbf{F}^{\text{mj}} (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(k-2-i)} (t_\alpha \mathbf{P}_{\text{occ}} \bar{\mathbf{R}} \mathbf{P}_{\text{occ}}) \right], \end{aligned} \quad (\text{B6b})$$

$$\bar{\mathbf{L}}^{(k)} = \frac{1}{k} [\bar{\mathbf{L}}^{(k-1)} (t_\alpha \mathbf{F} \mathbf{P}_{\text{occ}}) + (t_\alpha \mathbf{P}_{\text{occ}} \mathbf{F})^{(k-1)} (t_\alpha \mathbf{P}_{\text{occ}} \bar{\mathbf{R}}^{\text{mj}} \mathbf{P}_{\text{occ}})]. \quad (\text{B6c})$$

The starting terms are $\bar{\mathbf{M}}^{(0)} = \mathbf{0}$, $\bar{\mathbf{K}}^{(0)} = \mathbf{0}$, and $\bar{\mathbf{L}}^{(0)} = \mathbf{0}$. Likewise $\underline{\mathbf{M}}$, $\underline{\mathbf{K}}$, and $\underline{\mathbf{L}}$ can be obtained by inserting \mathbf{P}_{virt} instead of \mathbf{P}_{occ} and $\underline{\mathbf{R}}$ instead of $\bar{\mathbf{R}}$.

APPENDIX C: CPSCF EQUATIONS

The density matrix-based CPSCF equations, which are used for the AO-based Z-vector method, can be obtained by differentiating the equation of motion for the one-particle density matrix,^{26,46} which is for a stationary state:

$$\mathbf{F}\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{F} = \mathbf{0}. \quad (\text{C1})$$

The derivative with respect to a perturbation \mathbf{x} is

$$\begin{aligned} & \underbrace{\mathbf{F}\mathbf{P}^{\text{x}}\mathbf{S} - \mathbf{S}\mathbf{P}^{\text{x}}\mathbf{F} + \mathbf{G}[\mathbf{P}^{\text{x}}]\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{G}[\mathbf{P}^{\text{x}}]}_{\underline{\mathbf{A}}\mathbf{P}^{\text{x}}} \\ & = \underbrace{\mathbf{S}\mathbf{P}\mathbf{F}^{(\text{x})} - \mathbf{F}^{(\text{x})}\mathbf{P}\mathbf{S} + \mathbf{S}^{\text{x}}\mathbf{P}\mathbf{F} - \mathbf{F}\mathbf{P}\mathbf{S}^{\text{x}}}_{\mathbf{b}^{\text{x}}}. \end{aligned} \quad (\text{C2})$$

All terms, that are dependent on \mathbf{P}^{x} are shifted to the left-hand side and the other terms to the right-hand side. The left-hand side can be abbreviated by $\underline{\mathbf{A}}\mathbf{P}^{\text{x}}$ and the right-hand side by \mathbf{b}^{x} .

The derivative of Eq. (C2) with respect to a perturbation \mathbf{y} is

$$\begin{aligned} & \mathbf{F}\mathbf{P}^{\text{xy}}\mathbf{S} - \mathbf{S}\mathbf{P}^{\text{xy}}\mathbf{F} + \mathbf{G}[\mathbf{P}^{\text{xy}}]\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{G}[\mathbf{P}^{\text{xy}}] \\ & = -\mathbf{F}^{\text{y}}\mathbf{P}^{\text{x}}\mathbf{S} - \mathbf{F}\mathbf{P}^{\text{x}}\mathbf{S}^{\text{y}} + \mathbf{S}^{\text{y}}\mathbf{P}^{\text{x}}\mathbf{F} + \mathbf{S}\mathbf{P}^{\text{x}}\mathbf{F}^{\text{y}} \\ & \quad - \mathbf{G}^{\text{y}}[\mathbf{P}^{\text{x}}]\mathbf{P}\mathbf{S} - \mathbf{G}[\mathbf{P}^{\text{x}}]\mathbf{P}^{\text{y}}\mathbf{S} - \mathbf{G}[\mathbf{P}^{\text{x}}]\mathbf{P}\mathbf{S}^{\text{y}} \\ & \quad + \mathbf{S}^{\text{y}}\mathbf{P}\mathbf{G}[\mathbf{P}^{\text{x}}] + \mathbf{S}\mathbf{P}^{\text{y}}\mathbf{G}[\mathbf{P}^{\text{x}}] \\ & \quad + \mathbf{S}\mathbf{P}\mathbf{G}^{\text{y}}[\mathbf{P}^{\text{x}}] + \mathbf{S}^{\text{y}}\mathbf{P}\mathbf{F}^{(\text{x})} + \mathbf{S}\mathbf{P}^{\text{y}}\mathbf{F}^{(\text{x})} + \mathbf{S}\mathbf{P}\mathbf{F}^{(\text{xy})} - \mathbf{F}^{(\text{xy})}\mathbf{P}\mathbf{S} \\ & \quad - \mathbf{F}^{(\text{x})}\mathbf{P}^{\text{y}}\mathbf{S} - \mathbf{F}^{(\text{x})}\mathbf{P}\mathbf{S}^{\text{y}} + \mathbf{S}^{\text{xy}}\mathbf{P}\mathbf{F} + \mathbf{S}^{\text{x}}\mathbf{P}^{\text{y}}\mathbf{F} + \mathbf{S}^{\text{x}}\mathbf{P}\mathbf{F}^{\text{y}} \\ & \quad - \mathbf{F}^{\text{y}}\mathbf{P}\mathbf{S}^{\text{x}} - \mathbf{F}\mathbf{P}^{\text{y}}\mathbf{S}^{\text{x}} - \mathbf{F}\mathbf{P}\mathbf{S}^{\text{xy}}. \end{aligned} \quad (\text{C3})$$

The left-hand side of this equation can be abbreviated by $\underline{\mathbf{A}}\mathbf{P}^{\text{xy}}$. Furthermore, the terms of the right-hand side that include \mathbf{P}^{x} are

$$\begin{aligned} \underline{\mathbf{A}}^{\text{y}}\mathbf{P}^{\text{x}} = & \mathbf{F}^{\text{y}}\mathbf{P}^{\text{x}}\mathbf{S} + \mathbf{F}\mathbf{P}^{\text{x}}\mathbf{S}^{\text{y}} - \mathbf{S}^{\text{y}}\mathbf{P}^{\text{x}}\mathbf{F} - \mathbf{S}\mathbf{P}^{\text{x}}\mathbf{F}^{\text{y}} \\ & + \mathbf{G}^{\text{y}}[\mathbf{P}^{\text{x}}]\mathbf{P}\mathbf{S} + \mathbf{G}[\mathbf{P}^{\text{x}}]\mathbf{P}^{\text{y}}\mathbf{S} \\ & + \mathbf{G}[\mathbf{P}^{\text{x}}]\mathbf{P}\mathbf{S}^{\text{y}} - \mathbf{S}^{\text{y}}\mathbf{P}\mathbf{G}[\mathbf{P}^{\text{x}}] - \mathbf{S}\mathbf{P}^{\text{y}}\mathbf{G}[\mathbf{P}^{\text{x}}] - \mathbf{S}\mathbf{P}\mathbf{G}^{\text{y}}[\mathbf{P}^{\text{x}}], \end{aligned} \quad (\text{C4})$$

and the other terms are abbreviated by

$$\begin{aligned} \mathbf{b}^{\text{xy}} = & \mathbf{S}^{\text{y}}\mathbf{P}\mathbf{F}^{(\text{x})} + \mathbf{S}\mathbf{P}^{\text{y}}\mathbf{F}^{(\text{x})} + \mathbf{S}\mathbf{P}\mathbf{F}^{(\text{xy})} \\ & - \mathbf{F}^{(\text{xy})}\mathbf{P}\mathbf{S} - \mathbf{F}^{(\text{x})}\mathbf{P}^{\text{y}}\mathbf{S} - \mathbf{F}^{(\text{x})}\mathbf{P}\mathbf{S}^{\text{y}} \\ & + \mathbf{S}^{\text{xy}}\mathbf{P}\mathbf{F} + \mathbf{S}^{\text{x}}\mathbf{P}^{\text{y}}\mathbf{F} + \mathbf{S}^{\text{x}}\mathbf{P}\mathbf{F}^{\text{y}} - \mathbf{F}^{\text{y}}\mathbf{P}\mathbf{S}^{\text{x}} \\ & - \mathbf{F}\mathbf{P}^{\text{y}}\mathbf{S}^{\text{x}} - \mathbf{F}\mathbf{P}\mathbf{S}^{\text{xy}}. \end{aligned} \quad (\text{C5})$$

APPENDIX D: Z-VECTOR EQUATIONS FOR THE SECOND DERIVATIVE BY USING DL-CPSCF

In the following, details of the employed D-CPSCF equations for avoiding the second derivative of the density matrix are given. First, the derivative of the density matrix can be split according to subspace projections:

$$\text{Tr}[\mathcal{P}\mathbf{P}^{\text{Bmj}}] = \text{Tr}[\mathcal{P}(\mathbf{P}_{\text{ov}}^{\text{Bmj}} + \mathbf{P}_{\text{vo}}^{\text{Bmj}}) + \mathcal{P}(\mathbf{P}_{\text{oo}}^{\text{Bmj}} + \mathbf{P}_{\text{vv}}^{\text{Bmj}})]. \quad (\text{D1})$$

As shown in Sec. II B 3, the term $\text{Tr}[\mathcal{P}\mathbf{P}^{\text{Bmj}}]$ in Eq. (18) can be substituted by $\text{Tr}[\mathbf{Z}^T(\mathbf{b}^{\text{Bmj}} - \underline{\mathbf{A}}^{\text{mj}}\mathbf{P}^{\text{B}})]$. This can also be applied to each subspace projection of the \mathbf{P}^{Bmj} matrix:

$$\begin{aligned} \text{Tr}[\mathcal{P}\mathbf{P}^{\text{Bmj}}] = & \text{Tr}[(\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T)(\mathbf{b}^{\text{Bmj}} - \underline{\mathbf{A}}^{\text{mj}}\mathbf{P}^{\text{B}}) \\ & + \mathcal{P}(\mathbf{P}_{\text{oo}}^{\text{Bmj}} + \mathbf{P}_{\text{vv}}^{\text{Bmj}})], \end{aligned} \quad (\text{D2})$$

where $\mathbf{P}_{\text{oo}}^{\text{Bmj}}$ and $\mathbf{P}_{\text{vv}}^{\text{Bmj}}$ can simply be obtained non-iteratively by using the idempotency relation ($\mathbf{P}\mathbf{S}\mathbf{P} = \mathbf{P}$).^{26,47} By differentiating the idempotency relation twice and projecting onto the respective subspaces, one can calculate them directly by

$$\begin{aligned} \mathbf{P}_{\text{oo}}^{\text{Bmj}} = & -(\mathbf{P}\mathbf{S}\mathbf{P}^{\text{B}}\mathbf{S}\mathbf{P}^{\text{mj}}\mathbf{S}\mathbf{P} + \mathbf{P}\mathbf{S}\mathbf{P}^{\text{mj}}\mathbf{S}^{\text{B}}\mathbf{P} + \mathbf{P}\mathbf{S}^{\text{B}}\mathbf{P}^{\text{mj}}\mathbf{S}\mathbf{P} \\ & + \mathbf{P}\mathbf{S}\mathbf{P}^{\text{mj}}\mathbf{S}\mathbf{P}^{\text{B}}\mathbf{S}\mathbf{P}) \end{aligned} \quad (\text{D3a})$$

$$\mathbf{P}_{\text{vv}}^{\text{Bmj}} = \mathbf{P}_{\text{oo}}^{\text{Bmj}} + \mathbf{P}^{\text{B}}\mathbf{S}\mathbf{P}^{\text{mj}} + \mathbf{P}^{\text{mj}}\mathbf{S}\mathbf{P}^{\text{B}} + \mathbf{P}^{\text{mj}}\mathbf{S}^{\text{B}}\mathbf{P} + \mathbf{P}\mathbf{S}^{\text{B}}\mathbf{P}^{\text{mj}}. \quad (\text{D3b})$$

For this reason, only the virtual-occupied and occupied-virtual parts need to be computed iteratively.

To obtain the Z-vector, one has to solve

$$\underline{\mathbf{A}}\mathbf{Z} = \mathcal{P}, \quad (\text{D4})$$

where $\underline{\mathbf{A}}$ is the symmetric, positive-definite Hessian.

The corresponding product can be calculated by the left-hand side in Eq. (C2) using \mathbf{Z} instead of \mathbf{P}^{x} :

$$\underline{\mathbf{A}}\mathbf{Z} = \mathbf{F}\mathbf{Z}\mathbf{S} - \mathbf{S}\mathbf{Z}\mathbf{F} + \mathbf{G}[\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}]\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{G}[\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}]. \quad (\text{D5})$$

Suitable expressions can be obtained by multiplying from the left with $(\mathbf{I} - \mathbf{S}\mathbf{P})$ and from the right with $(\mathbf{P}\mathbf{S})$, so that the following expression for the Z-vector results:

$$\mathcal{P}_{\text{vo}} = \mathbf{F}\mathbf{Z}_{\text{vo}}\mathbf{S} - \mathbf{S}\mathbf{Z}_{\text{vo}}\mathbf{F} + \mathbf{G}_{\text{vo}}[\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}]. \quad (\text{D6})$$

As shown by Beer and Ochsenfeld,²⁷ the DL-CPSCF method is an efficient way to solve this equation by using

$$\mathbf{Z}_{\text{vo}} = \sum_{\alpha=1}^{\tau} w_{\alpha} \bar{\mathbf{P}} (\mathcal{P}_{\text{vo}} - \mathbf{G}_{\text{vo}}[\mathbf{Z}_{\text{ov}} + \mathbf{Z}_{\text{vo}}]) \underline{\mathbf{P}}, \quad (\text{D7})$$

where \bar{P} is a virtual and \underline{P} an occupied pseudo-density (Eqs. (A4a) and (A4b)).

Moreover, the term in Eq. (D2) contains the derivative of the right-hand side of the D-CPSCF equations and the derivative of the symmetric, positive-definite Hessian (the general derivation of this equation is shown in Appendix C):

$$\begin{aligned} \mathbf{b}^{\text{Bm}_j} - \underline{\mathbf{A}}^{\text{m}_j} \mathbf{P}^{\text{B}} &= -\mathbf{F}^{(\text{Bm}_j)} \mathbf{P}^{\text{S}} + \text{SPF}^{(\text{Bm}_j)} - \mathbf{F}^{(\text{B})} \mathbf{P}^{\text{m}_j} \mathbf{S} \\ &\quad - \mathbf{F}^{\text{m}_j} \mathbf{P}^{\text{B}} \mathbf{S} - \mathbf{F}^{\text{m}_j} \mathbf{P}^{\text{S}} \mathbf{B} - \mathbf{F} \mathbf{P}^{\text{m}_j} \mathbf{S}^{\text{B}} + \mathbf{S}^{\text{B}} \mathbf{P}^{\text{m}_j} \mathbf{F} \\ &\quad + \mathbf{S}^{\text{B}} \mathbf{P} \mathbf{F}^{\text{m}_j} + \mathbf{S} \mathbf{P}^{\text{B}} \mathbf{F}^{\text{m}_j} + \mathbf{S} \mathbf{P}^{\text{m}_j} \mathbf{F}^{(\text{B})} - \mathbf{G}[\mathbf{P}^{\text{B}}] \mathbf{P}^{\text{m}_j} \mathbf{S} \\ &\quad + \mathbf{S} \mathbf{P}^{\text{m}_j} \mathbf{G}[\mathbf{P}^{\text{B}}], \end{aligned} \quad (\text{D8})$$

where the second derivative of the Fock matrix is defined by

$$\mathbf{F}^{(\text{Bm}_j)} = \mathbf{h}^{\text{Bm}_j} + \mathbf{G}^{\text{B}}[\mathbf{P}^{\text{m}_j}] + \mathbf{G}[\mathbf{P}_{\text{oo}}^{\text{Bm}_j} + \mathbf{P}_{\text{vv}}^{\text{Bm}_j}]. \quad (\text{D9})$$

APPENDIX E: Z-VECTOR EQUATIONS FOR THE FIRST DERIVATIVE BY USING DL-CPSCF

In Sec. II C, the abbreviated D-CPSCF equations were used to illustrate the basic approach for avoiding the calculation of \mathbf{P}^{B} by applying the AO-based Z-vector method. In the following, we provide the explicit definitions of the D-CPSCF equations. All the terms that include \mathbf{P}^{B} in Eqs. (11), (D8), (D3a), and (D3b) are combined to yield

$$\begin{aligned} \mathcal{O}^{\text{m}_j} \mathbf{P}^{\text{B}} &= \mathcal{P}^{\text{m}_j} \mathbf{P}^{\text{B}} + (\mathbf{F}^{\text{m}_j} (\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T) \mathbf{S} - \mathbf{S} (\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T) \mathbf{F}^{\text{m}_j} \\ &\quad + \mathbf{G}[-\mathbf{P}^{\text{m}_j} \mathbf{S} (\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T) + (\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T) \mathbf{S} \mathbf{P}^{\text{m}_j}]) \mathbf{P}^{\text{B}} \\ &\quad + (-2(\mathbf{S} \mathbf{P}^{\text{m}_j} \mathbf{S} \mathbf{P} (\mathbf{G}[-\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T] + \mathcal{P}) \mathbf{P}^{\text{S}} \\ &\quad + \mathbf{S} \mathbf{P} (\mathbf{G}[-\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T] + \mathcal{P}) \mathbf{P}^{\text{S}} \mathbf{P}^{\text{m}_j} \mathbf{S}) \\ &\quad + \mathbf{S} \mathbf{P}^{\text{m}_j} (\mathbf{G}[-\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T] + \mathcal{P}) \\ &\quad + (\mathbf{G}[-\mathbf{Z}_{\text{ov}}^T + \mathbf{Z}_{\text{vo}}^T] + \mathcal{P}) \mathbf{P}^{\text{m}_j} \mathbf{S}) \mathbf{P}^{\text{B}}, \end{aligned} \quad (\text{E1})$$

with

$$\begin{aligned} \mathcal{P}^{\text{m}_j} &= \bar{\mathbf{Y}}_1^{\text{m}_j} - \underline{\mathbf{Y}}_1^{\text{m}_j} + \mathbf{G}[\bar{\mathbf{Y}}_2^{\text{m}_j} + \underline{\mathbf{Y}}_2^{\text{m}_j}] + \bar{\mathbf{R}}^{\text{m}_j} e^{t_{\text{a}} \mathbf{P}_{\text{occ}} \mathbf{F}} \\ &\quad + \bar{\mathbf{R}} (e^{t_{\text{a}} \mathbf{P}_{\text{occ}} \mathbf{F}})^{\text{m}_j} - \underline{\mathbf{R}}^{\text{m}_j} e^{-t_{\text{a}} \mathbf{P}_{\text{virt}} \mathbf{F}} - \underline{\mathbf{R}} (e^{-t_{\text{a}} \mathbf{P}_{\text{virt}} \mathbf{F}})^{\text{m}_j}. \end{aligned} \quad (\text{E2})$$

Furthermore, \mathbf{P}^{B} is split in its subspace projections:

$$\text{Tr}[\mathcal{O}^{\text{m}_j} \mathbf{P}^{\text{B}}] = \text{Tr}[\mathcal{O}^{\text{m}_j} (\mathbf{P}_{\text{ov}}^{\text{B}} + \mathbf{P}_{\text{vo}}^{\text{B}} + \mathbf{P}_{\text{oo}}^{\text{B}} + \mathbf{P}_{\text{vv}}^{\text{B}})]. \quad (\text{E3})$$

In Sec. II C, we have shown, that $\text{Tr}[\mathcal{O}^{\text{m}_j} \mathbf{P}^{\text{B}}]$ can be substituted by $\text{Tr}[(\mathbf{Z}^T)^{\text{m}_j} \mathbf{b}^{\text{B}}]$, which can also be employed for each subspace projection:

$$\text{Tr}[\mathcal{O}^{\text{m}_j} \mathbf{P}^{\text{B}}] = \text{Tr}[(\mathbf{Z}_{\text{ov}}^T)^{\text{m}_j} + (\mathbf{Z}_{\text{vo}}^T)^{\text{m}_j} \mathbf{b}^{\text{B}} + \mathcal{O}^{\text{m}_j} \mathbf{P}_{\text{oo}}^{\text{B}}], \quad (\text{E4})$$

where the virtual-virtual part is zero and the occupied-occupied part is simply $\mathbf{P}_{\text{oo}}^{\text{B}} = -\mathbf{P}^{\text{S}} \mathbf{B} \mathbf{P}$. Therefore, only the virtual-occupied and occupied-virtual parts need to be obtained iteratively.

The Z-vector equations to be solved are

$$\begin{aligned} \underline{\mathbf{A}} \mathbf{Z}^{\text{m}_j} &= \mathcal{O}^{\text{m}_j} = \mathbf{F} \mathbf{Z}^{\text{m}_j} \mathbf{S} - \mathbf{S} \mathbf{Z}^{\text{m}_j} \mathbf{F} + \mathbf{G}[\mathbf{Z}_{\text{ov}}^{\text{m}_j} + \mathbf{Z}_{\text{vo}}^{\text{m}_j}] \mathbf{P}^{\text{S}} \\ &\quad - \text{SPG}[\mathbf{Z}_{\text{ov}}^{\text{m}_j} + \mathbf{Z}_{\text{vo}}^{\text{m}_j}]. \end{aligned} \quad (\text{E5})$$

Again, one projects onto the virtual-occupied subspace

$$\mathcal{O}_{\text{vo}}^{\text{m}_j} = \mathbf{F} \mathbf{Z}_{\text{vo}}^{\text{m}_j} \mathbf{S} - \mathbf{S} \mathbf{Z}_{\text{vo}}^{\text{m}_j} \mathbf{F} + \mathbf{G}_{\text{vo}}[\mathbf{Z}_{\text{ov}}^{\text{m}_j} + \mathbf{Z}_{\text{vo}}^{\text{m}_j}], \quad (\text{E6})$$

and, as before, uses the DL-CPSCF method to obtain the Z-vector

$$\mathbf{Z}_{\text{vo}}^{\text{m}_j} = \sum_{\alpha=1}^{\tau} w_{\alpha} \bar{P} (\mathcal{O}_{\text{vo}}^{\text{m}_j} - \mathbf{G}_{\text{vo}}[\mathbf{Z}_{\text{ov}}^{\text{m}_j} + \mathbf{Z}_{\text{vo}}^{\text{m}_j}]) \underline{P}. \quad (\text{E7})$$

Furthermore, the right-hand side \mathbf{b}^{B} of the D-CPSCF equations, is defined as

$$\mathbf{b}^{\text{B}} = -\mathbf{F} \mathbf{P}^{\text{S}} \mathbf{B} + \mathbf{S}^{\text{B}} \mathbf{P} \mathbf{F} - \mathbf{F}^{(\text{B})} \mathbf{P}^{\text{S}} + \text{SPF}^{(\text{B})}. \quad (\text{E8})$$

Here, the Fock matrix derivative $\mathbf{F}^{(\text{B})}$ contains the following terms:

$$\mathbf{F}^{(\text{B})} = \mathbf{h}^{\text{B}} + \mathbf{G}^{\text{B}}[\mathbf{P}] - \mathbf{G}[\mathbf{P}^{\text{S}} \mathbf{B} \mathbf{P}], \quad (\text{E9})$$

where the last term describes the $\mathbf{G}[\mathbf{P}_{\text{oo}}^{\text{B}}]$ part.

- ¹R. Ditchfield, *Mol. Phys.* **27**, 789 (1974).
- ²K. Wolinski, J. F. Hinton, and P. Pulay, *J. Am. Chem. Soc.* **112**, 8251 (1990).
- ³M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, and H. Horn, *Theor. Chim. Acta* **83**, 455 (1992).
- ⁴R. M. Stevens, R. M. Pitzer, and W. N. Lipscomb, *J. Chem. Phys.* **38**, 550 (1963).
- ⁵G. Schreckenbach and T. Ziegler, *J. Phys. Chem.* **99**, 606 (1995).
- ⁶J. R. Cheeseman, G. W. Trucks, T. A. Keith, and M. J. Frisch, *J. Chem. Phys.* **104**, 5497 (1996).
- ⁷G. Rauhut, S. Puyeat, K. Wolinski, and P. Pulay, *J. Phys. Chem.* **100**, 6310 (1996).
- ⁸J. Gauss, *Chem. Phys. Lett.* **191**, 614 (1992).
- ⁹J. Gauss, *J. Chem. Phys.* **99**, 3629 (1993).
- ¹⁰J. Gauss and J. F. Stanton, *Adv. Chem. Phys.* **123**, 355 (2002).
- ¹¹J. Gauss and J. F. Stanton, *J. Chem. Phys.* **102**, 251 (1995).
- ¹²J. Gauss and J. F. Stanton, *J. Chem. Phys.* **104**, 2574 (1996).
- ¹³M. Kállay and J. Gauss, *J. Chem. Phys.* **120**, 6841 (2004).
- ¹⁴M. R. Hoffmann, D. J. Fox, J. F. Gaw, Y. Osamura, Y. Yamaguchi, R. S. Grev, G. Fitzgerald, H. F. Schaefer, P. J. Knowles, and N. C. Handy, *J. Chem. Phys.* **80**, 2660 (1984).
- ¹⁵K. Ruud, T. Helgaker, R. Kobayashi, P. Jørgensen, K. L. Bak, and H. J. A. Jensen, *J. Chem. Phys.* **100**, 8178 (1994).
- ¹⁶A. A. Auer, J. Gauss, and J. F. Stanton, *J. Chem. Phys.* **118**, 10407 (2003).
- ¹⁷S. Moon and D. A. Case, *J. Comput. Chem.* **27**, 825 (2006).
- ¹⁸D. Flai, M. Beer, and C. Ochsenfeld, *J. Chem. Theory Comput.* **8**, 2260 (2012).
- ¹⁹L. Olsson and D. Cremer, *J. Chem. Phys.* **105**, 8995 (1996).
- ²⁰F. London, *J. Phys. Radium* **8**, 397 (1937).
- ²¹W. Kutzelnigg, *Isr. J. Chem.* **19**, 193 (1980).
- ²²M. Schindler and W. Kutzelnigg, *J. Chem. Phys.* **76**, 1919 (1982).
- ²³J. Gauss, in *Modern Methods and Algorithms of Quantum Chemistry*, edited by J. Grotendorst (John von Neumann Institute for Computing, Jülich, 2000), Vol. 3, pp. 541–592.
- ²⁴H. F. Hameka, *Mol. Phys.* **1**, 203 (1958).
- ²⁵C. Ochsenfeld, J. Kussmann, and F. Koziol, *Angew. Chem., Int. Ed.* **43**, 4485 (2004).

174104-15 M. Maurer and C. Ochsenfeld

J. Chem. Phys. **138**, 174104 (2013)

- ²⁶J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.* **127**, 054103 (2007).
- ²⁷M. Beer and C. Ochsenfeld, *J. Chem. Phys.* **128**, 221102 (2008).
- ²⁸M. Beer, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.* **134**, 074102 (2011).
- ²⁹E. D. Simandiras, R. D. Amos, and N. C. Handy, *Chem. Phys.* **114**, 9 (1987).
- ³⁰M. Kollwitz, M. Häser, and J. Gauss, *J. Chem. Phys.* **108**, 8295 (1998).
- ³¹M. Kollwitz and J. Gauss, *Chem. Phys. Lett.* **260**, 639 (1996).
- ³²J. Gauss and H.-J. Werner, *Phys. Chem. Chem. Phys.* **2**, 2083 (2000).
- ³³S. Loibl and M. Schütz, *J. Chem. Phys.* **137**, 084107 (2012).
- ³⁴J. Almlöf, *Chem. Phys. Lett.* **181**, 319 (1991).
- ³⁵M. Häser and J. Almlöf, *J. Chem. Phys.* **96**, 489 (1992).
- ³⁶M. Häser, *Theor. Chim. Acta* **87**, 147 (1993).
- ³⁷S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, *J. Chem. Phys.* **136**, 144107 (2012).
- ³⁸S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.* **138**, 014101 (2013).
- ³⁹N. C. Handy and H. F. Schaefer III, *J. Chem. Phys.* **81**, 5031 (1984).
- ⁴⁰Y. Yamaguchi, J. D. Goddard, Y. Osamura, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory* (Oxford University Press Inc, 1994).
- ⁴¹S. Schweizer, B. Doser, and C. Ochsenfeld, *J. Chem. Phys.* **128**, 154101 (2008).
- ⁴²P. R. Surján, *Chem. Phys. Lett.* **406**, 318 (2005).
- ⁴³P. Y. Ayala and G. E. Scuseria, *J. Chem. Phys.* **110**, 3660 (1999).
- ⁴⁴C. Ochsenfeld and M. Head-Gordon, *Chem. Phys. Lett.* **270**, 399 (1997).
- ⁴⁵TURBOMOLE, Version 6.3, a quantum chemical program package written by R. Ahlrichs, M. K. Armbruster, R. A. Bachorz, M. Bär, H. P. Baron, R. Bauernschmitt, F. A. Bischoff, S. Bärker, N. Crawford, P. Deglmann, F. Della Sala, M. Diedenhofen, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, A. Glöck, F. Haase, M. Häser, C. Hättig, A. Hellweg, S. Höfener, H. Horn, C. Huber, U. Huniar, M. Kattannek, W. Klopper, A. Köhn, C. Kölmel, M. Kollwitz, K. May, P. Nava, C. Ochsenfeld, H. Ohm, M. Pabst, H. Patzelt, D. Rappoport, O. Rubner, A. Schäfer, U. Schneider, M. Sierka, D. P. Tew, O. Treutler, B. Unterreiner, M. von Arnim, F. Weigend, P. Weis, H. Weiss, N. Winter with contributions from M. Dolg, J. Gauss, C. van Wüllen, S. Brode, and H. Schiffer, see <http://www.turbomole.com>.
- ⁴⁶P. A. M. Dirac, *Proc. Cambridge Philos. Soc.* **27**, 240 (1931).
- ⁴⁷R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd ed. (Academic Press Inc, 1989).

**5.2 Paper II: "Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 levels",
D. Flaig, M. Maurer, M. Hanni,
K. Braunger, L. Kick, M. Thubauville,
C. Ochsenfeld,
J. Chem. Theory Comput., **10**, 572 (2014)**

Reproduced with permission from *J. Chem. Theory Comput.*, **10**, 572 (2014).
Copyright 2014 American Chemical Society.

Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels

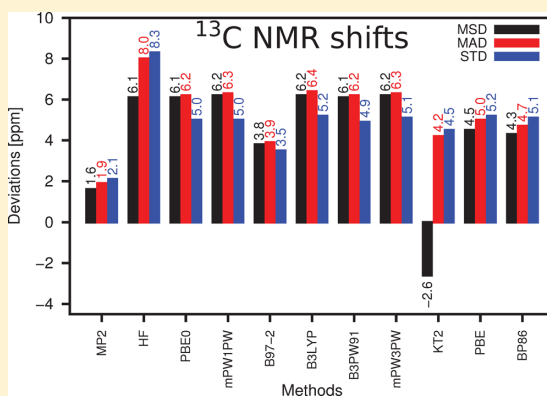
Denis Flaig, Marina Maurer, Matti Hanni, Katharina Braunger, Leonhard Kick, Matthias Thubauville, and Christian Ochsenfeld*

Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany

Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: An extensive study of error distributions for calculating hydrogen and carbon NMR chemical shifts at Hartree–Fock (HF), density functional theory (DFT), and Møller–Plesset second-order perturbation theory (MP2) levels is presented. Our investigation employs accurate CCSD(T)/cc-pVQZ calculations for providing reference data for 48 hydrogen and 40 carbon nuclei within an extended set of chemical compounds covering a broad range of the NMR scale with high relevance to chemical applications, especially in organic chemistry. Besides the approximations of HF, a variety of DFT functionals, and conventional MP2, we also present results with respect to a spin component-scaled MP2 (GIAO-SCS-MP2) approach. For each method, the accuracy is analyzed in detail for various basis sets, allowing identification of efficient combinations of method and basis set approximations.



1. INTRODUCTION

Over the past decades, much progress has been made in the ab initio calculation of nuclear-magnetic resonance (NMR) data of molecular systems (for an overview see, e.g., refs 1–5). Therefore, quantum-chemical calculations have become highly important tools in the often difficult assignment of experimental NMR spectra and can be performed today at various levels of theory ranging from Hartree–Fock (HF)^{6–8} or density-functional theory (DFT)⁹ to wave function-based correlation methods like, e.g., Møller–Plesset second-order perturbation theory (MP2)^{10–12} and coupled-cluster (CC) methods such as CC singles doubles (CCSD)¹³ or CCSD including perturbative triples (CCSD(T)).¹⁴ While in particular the high-accuracy schemes such as CCSD or CCSD(T) are confined to rather small molecules, also the more approximate quantum-chemical schemes are severely hampered in their applicability to larger molecules by the steep scaling of the computational effort with molecular size. To overcome these limitations, linear-scaling methods for calculating NMR shieldings have been devised at the HF and DFT levels (instead of their conventional cubic scaling), opening the way to calculate molecules with more than 1000 atoms on workstation computers.^{15–20} In addition, often just a few NMR shieldings are of interest (e.g., for a molecule within a solvent environment, where the shieldings of each solvent

molecule are irrelevant), so that most recently nuclei-selected NMR methods have been introduced that allow exploitation of the locality of the perturbation and reduction of the computational effort even beyond linear to sublinear, $O(M^0)$, i.e., asymptotically independent of molecular size.²¹ Also, for the simplest wave function-based correlation theory, MP2, such linear- or sublinear-scaling methods (instead of the conventional $O(M^5)$ scaling) have recently been formulated.²² While the latter schemes are still in a pilot-implementation stage, much of the method progress made for the calculation of MP2 energies for large molecules²³ is expected to be transferable.

Besides progress in developing fast methods, it is crucial for an efficient calculation of NMR shieldings to establish by extensive benchmarks the reliability of both the various approximation methods for solving the Schrödinger equation and the commonly used incomplete basis sets. Here, extensive and reliable information on the error distribution of the existing methods is required, which is the goal of our present work. In our present work, we benchmark NMR data, where we build upon earlier studies of the accuracy of NMR shieldings calculations^{24–40} and aim to go beyond the former studies in the following aspects: (1) We enlarge an earlier test set used by

Received: September 2, 2013

Published: January 23, 2014



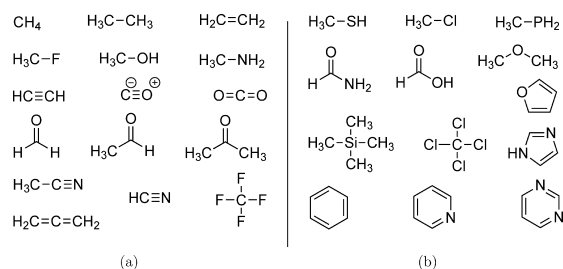


Figure 1. Molecular benchmark set of the current study: (a) Original molecular structures of ref 30. (b) Additional molecular structures of the present work.

Gauss and co-workers^{24,30} (see Figure 1a) by adding new molecules (see Figure 1b) with structures optimized at the CCSD(T)/cc-pVTZ level and with high relevance to organic chemistry (or biochemistry), covering a broad range of the NMR scale for proton and carbon shifts. In detail, the compounds benzene, furan, imidazole, pyridine, and pyrimidine were added to the original set representing aromatic systems (for example, their derivatives are typically comprised of proteins or nucleic acids etc.). Dimethyl ether, formic acid, and formamide serve as representatives for ethers, carboxylic acids, and carboxamides, respectively. As a further important system, the standard reference molecule tetramethylsilane (TMS) was included in the selection, allowing for calculations of standard relative shieldings for hydrogen and carbon nuclei. Additionally, the compounds CH₃PH₂, CH₃SH, CCl₄, and CH₃Cl are employed for considering further hetero atoms. (2) Besides carbon (or hetero) shieldings, we also consider hydrogen shieldings. (3) Our error analysis is based on rather accurate reference calculations at the CCSD(T)/cc-pVQZ level. (4) In addition, MP2 as well as spin-component-scaled-(SCS)-MP2 results are provided. MP2 theory represents the most cost-effective wave-function-based correlation theory and is known to provide often reliable NMR chemical shifts. (5) Results for a broad selection of different DFT functionals are presented. (6) The accuracy is analyzed for a variety of basis sets.

While for the original set of structures introduced by Gauss and co-workers also highly accurate gas phase NMR experimental measurements are available,⁴¹ our present study aims to extend the test set also to molecules for which no gas phase experiments are available. Therefore, we restrain ourselves to compare the accuracies to the most reliable theoretical NMR data available as computed at the CCSD(T)/cc-pVQZ level. In this way, influences of the structure, vibrations, etc. are eliminated, and we get the information on how to obtain the most reliable theoretical data in a most cost-efficient way (method/basis).

After describing some methodological aspects, we focus first on the accuracy of various quantum-chemical methods by largely eliminating basis set influences in using a large basis. Then cost-efficient pathways are discussed by studying the accuracy of a variety of smaller and medium-sized basis sets.

2. METHODOLOGICAL ASPECTS

The global minimum structures for all molecules of the benchmark set (see Figure 1) were obtained at the CCSD(T)/cc-pVTZ⁴² level by the program package CFOUR⁴³ and are available via the Web site <http://www.cup.uni-muenchen.de/p/ochsenfeld/download.html>.

The GIAO-(SCS)-MP2 NMR calculations were performed with a development version of the program package Q-Chem⁴⁴ based on an AO-MP2 NMR implementation.²² The DFT calculations were performed with the newly developed QM package FermiONS++,⁴⁵ which includes DFT functionals from the Libxc database.⁴⁶ The selection of DFT methods comprises hybrid GGA-functionals PBE0,⁴⁷ mPW1PW,⁴⁸ B97-2,⁴⁹ B3LYP,^{50,51} B3PW91,⁵² mPW3PW,⁴⁸ and pure GGA functionals BP86,⁵³ PBE,⁵⁴ and KT2.⁵⁵ The reference calculations at the CCSD(T) level were performed with the program package CFOUR.⁴³

The employed basis sets (STO-3G,^{56,57} 3-21G,^{58,59} 6-31G**,^{60,61} 6-311G**,⁶² pcS-0, pcS-1, pcS-2,³² def2-SVP, def2-TZVP,⁶³ tz2p, qz2p,^{30,64} cc-pVDZ, cc-pVTZ, cc-pVQZ⁴²) were partly transferred from the basis set exchange database.^{65,66}

3. ACCURACIES OF VARIOUS QUANTUM-CHEMICAL APPROXIMATIONS

In the following, we first aim at eliminating basis set influences by using a large basis set (cc-pVQZ⁴²) and focus on the expected errors of HF, different DFT functionals, and MP2 as compared to the most reliable data obtained at the CCSD(T) level. By comparing to the rather accurate theoretical results at the CCSD(T) level, we focus on the accuracies for the same structure and avoid influences of conformational changes (in particular molecular vibrations) or the chemical environment (e.g., a host system or solvent etc.). For a comparison to experimental data, see refs 24, 30, and 41.

Figures 2 and 3 summarize the effective errors for all tested methods for hydrogen and carbon nuclei, respectively. The

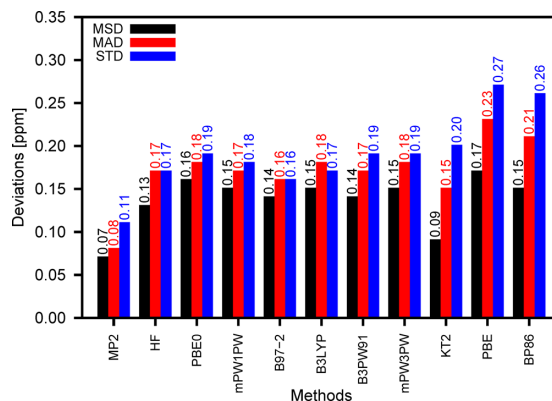


Figure 2. MSD and MAD with respect to TMS and the STD (in ppm) for ¹H NMR shifts at MP2, DFT, and HF levels with respect to CCSD(T) results (basis set cc-pVQZ and CCSD(T)/cc-pVTZ geometries; the GIAO approach is always employed).

explicit values of the underlying NMR shifts for all considered nuclei and all levels of theory are provided in the Supporting Information. For evaluating the accuracy of methods in calculating NMR shifts (relative shieldings of nuclei *A* with respect to the reference nucleus, e.g., in TMS: $\delta_A = \sigma_{\text{TMS}} - \sigma_A$), three distinctive error criteria are listed: the mean signed deviation (MSD), the mean absolute deviation (MAD), and the standard deviation (STD), which are computed as follows over *N* nuclei:

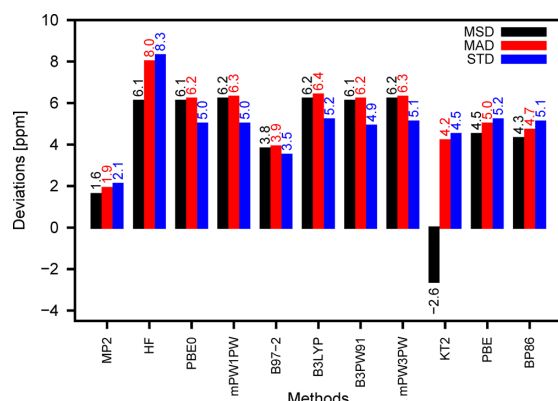


Figure 3. MSD and MAD with respect to TMS and the STD (in ppm) for ^{13}C NMR shifts at MP2, DFT, and HF levels with respect to CCSD(T) results (basis set cc-pVQZ and CCSD(T)/cc-pVTZ geometries; the GIAO approach is always employed).

$$\text{MSD} = \frac{1}{N} \sum_A \overbrace{\delta_A - \delta_A^{\text{CCSD(T)/cc-pVQZ}}}^{\Delta\delta_A}$$

$$\text{MAD} = \frac{1}{N} \sum_A |\Delta\delta_A|$$

$$\text{STD} = \sqrt{\frac{1}{N-1} \sum_A (\text{MSD} - \Delta\delta_A)^2}$$

Here, the standard deviation STD is invariant with respect to the analogous formulation with absolute shielding values σ instead of relative shifts δ (the MSD then also needs to be calculated over σ). Thus, the STD is independent of the selection of the reference nucleus, in contrast to the MSD or MAD, and therefore represents the most meaningful general criterion for judging the accuracy of NMR shifts. Because of the common use of TMS as the reference compound, the MSD and MAD for shifts with respect to TMS are also provided.

The data plotted in Figures 2 and 3 indicate that the MP2 method achieves the lowest standard deviations of 0.11/2.1 ppm for $^1\text{H}/^{13}\text{C}$ nuclei with respect to the CCSD(T) reference. In contrast, HF and DFT results range between 0.16/3.5 and

0.27/8.3 ppm for $^1\text{H}/^{13}\text{C}$. The order according to the accuracies differs to some extent for ^1H and ^{13}C nuclei: For ^1H shifts, HF proves to be on the same level of accuracy as compared to DFT, while for ^{13}C shifts all DFT functionals are better by more than 3.1 ppm. For both ^1H and ^{13}C shifts, the functional B97-2⁴⁹ performs best, followed by the functional KT2⁵⁵ in the case of ^{13}C shifts.

The good performance of B97-2 and KT2 is also reflected by the MAD values of shifts with respect to TMS. Although in general the MSD could be regarded to be a less meaningful error criterion for calculating relative shieldings (shifts), it shows an interesting behavior: The KT2 functional is the only method that systematically underestimates the carbon shifts calculated with respect to TMS (negative MSD), whereas all the other methods overestimate the standard shift (positive MSDs).

The Supporting Information gives further insights into the error distribution and the relationship to chemical structure. Generally, one can conclude that chemically similar structures (small absolute values of shifts) are better described than more different structures. Thus, if TMS is chosen as the reference compound, the problematic cases accumulate mainly in the regime of aromatic and carbonylic groups. Moreover, calculating carbon shifts for the compounds CO, CCl_4 , and the allene CH_2CCH_2 is shown to be especially demanding for the DFT methods. An inverse effect is observed, if for instance CO is chosen as the reference compound: now the sp^3 hybridized structures are the most deviating cases (see Section 15 of the Supporting Information and section 6 for a more detailed discussion).

4. BASIS SET INFLUENCES

While the use of the rather large basis set cc-pVQZ⁴² largely eliminates basis set effects for the data presented above, we focus in the following on the influence of basis set deficiencies. Clearly, the size of the basis set plays a central role for the computing time and may effectively limit the size of treatable systems. Therefore, the STD is listed in Tables 1 and 2 for hydrogen and carbon shifts as computed using a variety of basis sets for all different methods employed above.

Similar to the discussion above, the STD for all methods and basis sets is referenced to CCSD(T)/cc-pVQZ data. Thus, the

Table 1. Standard Deviation (STD) of H Shifts for Different Basis Sets and Methods with Respect to the Reference Calculation CCSD(T)/cc-pVQZ

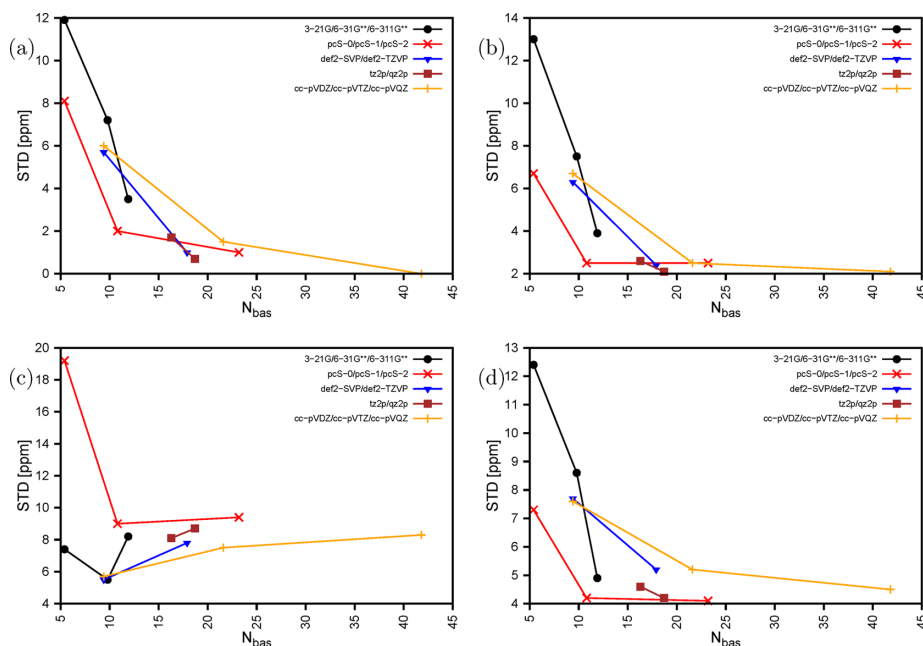
basis set	\bar{N}_{bas}^a	CCSD(T)	MP2	HF	PBE0	mPW1PW	B97-2	B3LYP	B3PW91	mPW3PW	KT2	PBE	BP86
STO-3G	3.0	0.77	0.79	1.02	0.86	0.86	0.83	0.85	0.85	0.85	0.77	0.82	0.81
3-21G	5.4	0.56	0.50	0.65	0.62	0.61	0.60	0.62	0.61	0.61	0.62	0.66	0.64
6-31G**	9.8	0.19	0.17	0.22	0.23	0.22	0.21	0.23	0.23	0.23	0.26	0.31	0.30
6-311G**	11.9	0.18	0.17	0.21	0.23	0.22	0.21	0.23	0.23	0.23	0.25	0.31	0.30
pcS-0	5.4	0.47	0.41	0.66	0.54	0.53	0.51	0.52	0.52	0.52	0.46	0.54	0.52
pcS-1	10.8	0.16	0.14	0.21	0.17	0.17	0.16	0.18	0.18	0.18	0.21	0.25	0.24
pcS-2	23.2	0.03	0.10	0.19	0.17	0.16	0.15	0.15	0.17	0.17	0.18	0.25	0.24
def2-SVP	9.4	0.23	0.21	0.22	0.21	0.21	0.20	0.22	0.21	0.22	0.25	0.29	0.28
def2-TZVP	17.9	0.11	0.13	0.18	0.23	0.22	0.20	0.23	0.23	0.23	0.26	0.32	0.31
tz2p	16.3	0.12	0.12	0.17	0.19	0.18	0.17	0.19	0.19	0.19	0.22	0.28	0.27
qz2p	18.7	0.09	0.10	0.17	0.19	0.18	0.16	0.18	0.19	0.19	0.22	0.28	0.26
cc-pVDZ	9.4	0.25	0.23	0.23	0.27	0.26	0.26	0.28	0.27	0.27	0.31	0.36	0.35
cc-pVTZ	21.6	0.07	0.11	0.16	0.19	0.18	0.16	0.18	0.19	0.19	0.22	0.28	0.27
cc-pVQZ	41.8	0.00	0.11	0.17	0.19	0.18	0.16	0.17	0.19	0.19	0.20	0.27	0.26

^aAverage number of basis functions per atom (determined for the molecular benchmark set).

Table 2. Standard Deviation (STD) of C Shifts for Different Basis Sets and Methods with Respect to the Reference Calculation CCSD(T)/cc-pVQZ

basis set	\bar{N}_{bas}^a	CCSD(T)	CCSD(T) ^b	MP2	MP2 ^b	HF	PBE0	mPW1PW	B97-2	B3LYP	B3PW91	mPW3PW	KT2	PBE	BP86
STO-3G	3.0	25.8	8.6	27.0	9.9	16.5	20.3	20.4	21.1	21.0	20.8	20.7	23.6	22.2	22.4
3-21G	5.4	11.9	4.4	13.0	6.4	7.4	7.6	7.5	8.6	8.2	7.9	7.9	12.4	10.0	10.1
6-31G**	9.8	7.2	2.2	7.5	2.7	5.5	3.3	3.4	4.4	4.5	3.8	3.8	8.6	5.9	6.2
6-311G**	11.9	3.5	1.8	3.9	2.9	8.2	3.4	3.3	2.5	3.5	3.3	3.4	4.9	3.7	3.7
pcS-0	5.4	8.1	6.8	6.7	9.4	19.2	11.2	11.1	9.8	10.4	10.7	10.8	7.3	9.6	9.4
pcS-1	10.8	2.0	1.4	2.5	2.6	9.0	4.6	4.5	3.2	4.7	4.4	4.5	4.2	4.5	4.3
pcS-2	23.2	1.0	0.5	2.5	2.2	9.4	6.2	6.2	4.5	6.4	6.1	6.2	4.1	6.0	5.8
def2-SVP	9.4	5.7	2.8	6.3	3.4	5.5	3.5	3.6	4.2	4.4	4.0	4.0	7.7	5.7	5.9
def2-TZVP	17.9	1.0	0.4	2.4	2.2	7.8	4.2	4.1	3.2	4.4	4.1	4.3	5.2	4.8	4.7
tz2p	16.3	1.7	0.8	2.6	2.2	8.1	3.9	3.8	2.7	4.0	3.7	3.9	4.6	4.2	4.1
qz2p	18.7	0.7	0.8	2.1	2.3	8.7	4.9	4.9	3.4	5.2	4.8	5.0	4.2	4.9	4.8
cc-pVDZ	9.4	6.0	2.9	6.7	3.5	5.7	3.8	3.8	4.4	4.6	4.1	4.2	7.6	5.8	6.0
cc-pVTZ	21.6	1.5	0.5	2.5	2.1	7.5	3.8	3.7	2.8	4.1	3.7	3.8	5.2	4.5	4.4
cc-pVQZ	41.8	0.0	0.0	2.1	2.1	8.3	5.0	5.0	3.5	5.2	4.9	5.1	4.5	5.2	5.1

^aAverage number of basis functions per atom (determined for the molecular benchmark set). ^bThe HF part of the C shifts is calculated with the basis set cc-pVQZ (see Section 5 for a detailed discussion).

**Figure 4.** Dependence of the STD for C shifts with respect to the reference calculation CCSD(T)/cc-pVQZ on the mean number of basis functions per atom N_{bas} for (a) CCSD(T), (b) MP2, (c) HF, and (d) KT2. Here, five different basis set series are distinguished.

first column of Tables 1 and 2 lists deviations for CCSD(T) calculations caused by smaller basis sets, while the other columns list deviations caused by the interplay of method and basis set approximations (in addition, by referring to the results obtained for a cc-pVQZ basis within the respective method, the pure effects of smaller basis sets for each method can be extracted; see the Supporting Information). The STD for the cc-pVQZ basis visualized in Figure 3 is listed again in the last row. The size of the basis sets is specified by a single measure, that is, the average number \bar{N}_{bas} of basis functions per atom computed over all atoms of the molecular benchmark set (the details for the basis set composition are provided in the Supporting Information in the form of contraction schemes).

The data for hydrogen shieldings in Table 1 show only weak influences by basis set changes (only for the rather poor basis sets STO-3G, 3-21G, and pcS-0, larger changes occur). Therefore, we focus in the following on carbon NMR shieldings as listed in Table 2: To begin with, the most pronounced increase of deviations for all methods occurs below the levels of double- ζ basis sets with polarization functions (DZP), e.g., 6-31G**, pcS-1 (as one might expect). The smallest basis set usable for wave-function-based correlation methods (CCSD(T), MP2) appears to be the pcS-1 basis introduced by Jensen.³² Here, the wave-function-based correlation methods reach a strikingly good performance as compared to DFT with a particularly good compromise between accuracy and cost: the

average number of basis functions per atom for the present benchmark set, abbreviated by \bar{N}_{bas} , is 10.8. Also for Hartree–Fock and the various DFT functionals, the pcS-1 basis set clearly comes closest to the corresponding STDs obtained with the cc-pVQZ basis set (among the basis sets of similar size 6-31G**, pcS-1, def2-SVP, or cc-pVDZ). A further analysis (see the Supporting Information) shows that standard deviations of only 1.4–1.6 ppm are reached for HF and DFT compared to the results obtained with the respective method at the cc-pVQZ level (as compared to 3.9–5.9 ppm, e.g., for def2-SVP).

For higher accuracies than obtained with the cost-effective pcS-1 basis, the data in Table 2 indicate the very good performance of def2-TZVP⁶³ and, in particular, qz2p basis sets^{30,64} with a size \bar{N}_{bas} of 17.9 and 18.7, respectively. The latter is only slightly larger than the def2-TZVP basis but still noticeably better performing for the calculation of NMR chemical shifts. These basis sets are clearly a much better high-accuracy compromise than the more than twice as large cc-pVQZ basis⁴² ($\bar{N}_{\text{bas}} = 41.8$).

While the error reduces consistently with the size of the basis set for the wave-function-based correlation methods CCSD(T) and MP2, as well as the KT2 functional, deviating behavior can be observed for the Hartree–Fock method and other considered DFT functionals. Here, the data indicate that method and basis set errors to some extent cancel statistically, especially in the regime of DZP quality (while of course there is no guarantee that for a specific nucleus an error cancellation occurs). The different behavior with increasing basis set size is visualized in Figure 4 for the four representative examples of CCSD(T), MP2, HF, and KT2. For the larger basis sets, the B97-2 functional performs best of the tested DFT functionals, followed by the KT2 functional.

More data are listed in the Supporting Information that indicates that the general aspects listed above are also reflected by other error criteria (e.g., mean absolute and maximum deviations).

5. ERROR DISTRIBUTION FOR SPIN COMPONENT-SCALED MP2 (GIAO-SCS-MP2) APPROACHES

In addition to the MP2 NMR shieldings, we also list data for our newly introduced GIAO-SCS-MP2 shieldings method, which employs scaling coefficients for the same- and opposite-spin components and for which details of the fitting procedure and extensive accuracy studies will be presented elsewhere.⁶⁷ The latter scheme follows the lines of scaled MP2 methods developed for ground state energetics by Grimme⁶⁸ and others (see, e.g., refs 69 and 70 or ref 71 for a recent review). Analogously, one can split the NMR shielding tensor at the MP2 level into the HF contribution and the opposite spin (OS) and same spin (SS) terms of the perturbative second-order correction:

$$\sigma_{\text{MP2}} = \sigma_{\text{HF}} + c_{\text{OS}} \sigma_{\text{OS}} + c_{\text{SS}} \sigma_{\text{SS}}$$

The scaling factors c_{OS} and c_{SS} have been optimized by a fitting procedure with respect to the NMR shieldings at the CCSD(T)/cc-pVQZ level for each single basis set. The fitting procedure itself employs only one-half of the nuclei of the total benchmark set, while for analyzing the error distribution we include all nuclei. Furthermore, in the fitting procedure, we allow for systematic deviations of (absolute) shieldings, because our focus is on relative shieldings (shifts) and include for this purpose a constant offset (denoted as MSD_{opt}):

$$\sigma_{\text{SCS/SOS-MP2}} + \text{MSD}_{\text{opt}} \stackrel{!}{=} \sigma_{\text{CCSD(T)}}$$

Table 3 lists the results for spin component-scaled MP2 approaches for NMR shifts with respect to the analogously computed TMS values:

Table 3. Standard Deviations (STD) of C and H Shifts for the GIAO-SCS-MP2 and GIAO-SOS-MP2 Methods Introduced in This Work Employing a Variety of Basis Sets^a

basis set	\bar{N}_{bas}^b	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2	B97-2
C shifts						
STO-3G	3.0	25.8	13.7	15.4	27.0	21.1
3-21G	5.4	11.9	5.1	7.4	13.0	8.6
6-31G**	9.8	7.2	2.4	4.4	7.5	4.4
6-311G**	11.9	3.5	2.6	3.7	3.9	2.5
pcS-0	5.4	8.1	6.5	6.7	6.7	9.8
pcS-1	10.8	2.0	2.2	2.9	2.5	3.2
pcS-0 ^c		6.8	3.1	3.4	9.4	
pcS-1 ^c		1.4	1.6	2.0	2.6	
H shifts						
pcS-0	5.4	0.47	0.36	0.45	0.41	0.51
pcS-1	10.8	0.16	0.14	0.18	0.14	0.16

^aAs the reference, CCSD(T)/cc-pVQZ results are used. For comparison, CCSD(T), MP2, and B97-2 results are again listed (see Table 2). More details for the GIAO-SCS-MP2 and GIAO-SOS-MP2 results will be presented elsewhere.⁶⁷ ^bAverage number of basis functions per atom (determined for the molecular benchmark set).

^cThe HF part of the C shifts is calculated with the basis set cc-pVQZ.

- In particular, for the smaller basis sets like STO-3G, 3-21G, and 6-31G**, the scaling of the MP2 values reduces the STDs strongly. For example, the STD using the 6-31G** basis is close to the STDs with larger basis sets like 6-311G** and pcS-1.

- The STD of GIAO-SCS-MP2/pcS-1 is close to the value of CCSD(T)/pcS-1.

The proposed GIAO-SCS-MP2 method introduced above employs the same basis set for the HF and the perturbative second-order correction terms. A pragmatic alternative and less common approach in striving for a good compromise between accuracy and cost would be to employ a larger basis for the HF term to reduce the HF error to a minimum: We tested here the cc-pVQZ basis set. The results for this additional approach are listed in Table 3: The STD for GIAO-SCS-MP2/pcS-0 reduces from 6.5 to 3.1 ppm and for pcS-1 from 2.2 to 1.6 ppm. Remarkably, the mixed basis set GIAO-SCS-MP2 approach results for the latter case in smaller STDs than obtained by conventional CCSD(T)/pcS-1 calculations. In contrast to the good benefit in case of the GIAO-SCS-MP2 method, an analogous mixed approach for nonscaled MP2 shows to be far less beneficial. Actually, for some cases one obtains even larger STDs. Table 2 gives an overview for the mixed basis set approach for all basis sets and also in combination with the CCSD(T) method. Especially, the improvements for CCSD(T) in combination with the basis sets pcS-2, def2-TZVP, and cc-pVTZ to STDs below 1 ppm are remarkable.

The relative benefits of the GIAO-SCS-MP2 method to the nonscaled MP2 method were also investigated for H shifts for the basis sets pcS-0 and pcS-1 (see Table 3): The STD of

GIAO-SCS-MP2/pcS-0 drops from 0.41 to 0.36 ppm, whereas the STD of GIAO-SCS-MP2/pcS-1 remains unchanged.

Further details for the GIAO-SCS-MP2 approach are given in the Supporting Information and in a future publication.⁶⁷

6. ERROR DISTRIBUTION FOR MOLECULAR SUBSETS

So far, we considered standard deviations for the complete molecular test set and reported mean absolute deviations with respect to TMS only. However, as concluded by several previous studies, the errors in calculating NMR shifts can be decreased substantially, if the shifts of the probed structure are computed with respect to a closely related structure. This refers to the concept of intermediate references,^{72,73} or multistandards (see, e.g., refs 74 and 75), where the total shift with respect to TMS, $\delta_{\text{TMS-sample}}^{\text{high/low}}$, is determined as the sum of a highly accurate theoretical or experimental shift value, $\delta_{\text{TMS-Int.ref.}}^{\text{high}}$, for the intermediate reference (or second standard compound in a multistandard approach) and an incremental shift, $\delta_{\text{Int.ref-sample}}^{\text{low}}$, computed at a lower level:

$$\delta_{\text{TMS-sample}}^{\text{high/low}} = \underbrace{\delta_{\text{TMS-Int.ref.}}^{\text{high}}}_{\sigma_{\text{TMS}}^{\text{high}}} - \underbrace{\delta_{\text{Int.ref.}}^{\text{high}}}_{\sigma_{\text{Int.ref.}}^{\text{high}}} + \underbrace{\delta_{\text{Int.ref-sample}}^{\text{low}}}_{\sigma_{\text{Int.ref.}}^{\text{low}}} - \underbrace{\delta_{\text{sample}}^{\text{low}}}_{\sigma_{\text{sample}}^{\text{low}}}$$

To appraise the possible accuracy gains by employing intermediate references (or multistandards), Figure 5 depicts

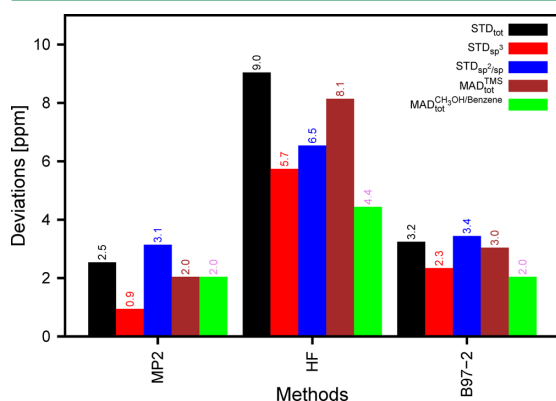


Figure 5. STDs (in ppm) for subsets of sp^3 or sp^2/sp hybridized carbon atoms and MADs with respect to TMS, and for an approach with two intermediate references (CH_3OH , benzene) at GIAO MP2, B97-2, and HF levels with respect to CCSD(T) results (basis set pcS-1 and CCSD(T)/cc-pVTZ geometries).

the standard deviations calculated for subsets of sp^3 or sp^2/sp hybridized carbon atoms for the example of the (well-performing) pcS-1 basis set and the MP2, HF, and B97-2 method. On the one hand, the figure indicates that electron correlation plays a major role within the sp^2/sp subset compared to the sp^3 subset (for all methods the $\text{STD}_{\text{sp}^2/\text{sp}}$ is larger than the STD_{sp^3}). Especially for MP2, the sp^3 subset is very well described with STDs of only 0.9 ppm. On the other hand, the results reflect the benefits from the separation into subsets and referring to a chemically related nucleus. For example, if all the sp^3 carbon atoms are referred to CH_3OH and the sp^2/sp carbon atoms are referred to benzene, an overall $\text{MAD}_{\text{tot}}^{\text{CH}_3\text{OH/benzene}}$ of 2.0/4.4/2.0 can be reached for MP2/HF/B97-2 (CH_3OH and benzene are chosen here as an example, motivated by refs 74 and 75). Here, the overall MAD for using

the two standards is computed as the average of the subset MADs weighted by the number of nuclei per subset (15 sp^3 and 25 sp^2/sp nuclei).

7. CONCLUSION

Within the present work, we have benchmarked the errors for calculating NMR shifts by MP2, HF, and DFT approaches based on a broad molecular set with large importance for computations of hydrogen and carbon NMR shifts in organic compounds. At the DZP level and above, the assigned standard deviations for hydrogen vary in between 0.10–0.23/0.16–0.23/0.15–0.36 ppm for MP2/HF/DFT and for carbon in between 2.1–7.4/5.5–9.4/2.5–8.6 ppm depending on the basis set. Here, among the DFT functionals, the considered hybrid-GGA functional B97-2 reaches rather constantly the lowest standard deviations. A further reduction of the standard deviation beyond the DFT level can be reached by MP2, where the study identifies the Jensen basis set pcS-1 as a remarkably well performing cost-efficient alternative to larger (TZP or QZP) basis sets. Furthermore, our new scaled MP2 approach for NMR shieldings is shown to be a helpful extension that can considerably reduce the deviation with respect to the CCSD(T)/cc-pVQZ reference values.

■ ASSOCIATED CONTENT

Supporting Information

Explicit NMR shift values for all considered nuclei and all levels of theory, additional error criteria, contraction schemes for all basis sets, and further details for the GIAO-SCS-MP2 approach. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: christian.ochsenfeld@cup.uni-muenchen.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. J. Kussmann (LMU) for support concerning calculations of NMR shielding tensors, J. Glänzer (student at LMU) for participating in geometry optimizations and NMR calculations during a practical course, and A. Lünser (LMU) for further helpful discussions. C.O. acknowledges financial support by the Volkswagen Stiftung within the funding initiative “New Conceptual Approaches to Modeling and Simulation of Complex Systems,” by the SFB 749 “Dynamik und Intermediate molekularer Transformationen” (DFG), and the DFG cluster of excellence EXC 114 “Center for Integrative Protein Science Munich” (CIPSM).

■ REFERENCES

- (1) Bühl, M.; Malkin, V. G.; Kaupp, M. *Calculation of NMR and EPR Parameters, Theory and Applications*; Wiley-VCH: Weinheim, Germany, 2004.
- (2) Gauss, J. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, 2000; Vol. 3, pp 541–592.
- (3) Vaara, J. *Phys. Chem. Chem. Phys.* **2007**, 9, 5399–5418.
- (4) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, 99, 293–352.
- (5) Kussmann, J.; Beer, M.; Ochsenfeld, C. *WIREs Comput. Mol. Sci.* **2013**, 3, 614–636.

- (6) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.
- (7) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- (8) Häser, M.; Ahlrichs, R.; Baron, H.; Weis, P.; Horn, H. *Theor. Chim. Acta* **1992**, *83*, 455–470.
- (9) Wolff, S. K.; Ziegler, T. *J. Chem. Phys.* **1998**, *109*, 895–905.
- (10) Gauss, J. *J. Chem. Phys.* **1993**, *99*, 3629–3643.
- (11) Kollwitz, M.; Gauss, J. *Chem. Phys. Lett.* **1996**, *260*, 639–646.
- (12) Kollwitz, M.; Häser, M.; Gauss, J. *J. Chem. Phys.* **1998**, *108*, 8295–8301.
- (13) Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **1995**, *103*, 3561–3577.
- (14) Gauss, J.; Stanton, J. F. *Chem. Phys. Lett.* **1997**, *276*, 70–77.
- (15) Ochsenfeld, C.; Head-Gordon, M. *Chem. Phys. Lett.* **1997**, *270*, 399–405.
- (16) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Chem., Int. Ed.* **2004**, *43*, 4485–4489.
- (17) Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2007**, *127*, 054103.
- (18) Beer, M.; Ochsenfeld, C. *J. Chem. Phys.* **2008**, *128*, 221102.
- (19) Larsen, H.; Helgaker, T.; Olsen, J.; Jorgensen, P. *J. Chem. Phys.* **2001**, *115*, 10344–10352.
- (20) Niklasson, A. M. N.; Challacombe, M. *Phys. Rev. Lett.* **2004**, *92*, 193001.
- (21) Beer, M.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2011**, *134*, 074102.
- (22) Maurer, M.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 174104.
- (23) Maurer, S. A.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 014101.
- (24) Gauss, J. *J. Chem. Phys.* **1993**, *99*, 3629–3643.
- (25) Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. *J. Chem. Phys.* **1996**, *104*, 5497–5509.
- (26) Olsson, L.; Cremer, D. *J. Chem. Phys.* **1996**, *105*, 8995–9006.
- (27) Adamo, C.; Barone, V. *Chem. Phys. Lett.* **1998**, *298*, 113–119.
- (28) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *J. Phys. Chem. A* **1999**, *103*, 7357–7363.
- (29) Sun, H.; Sanders, L. K.; Oldfield, E. *J. Am. Chem. Soc.* **2002**, *124*, 5486–5495.
- (30) Auer, A. A.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2003**, *118*, 10407–10417.
- (31) Manninen, P.; Vaara, J. *J. Comput. Chem.* **2006**, *27*, 434–445.
- (32) Jensen, F. *J. Chem. Theory Comput.* **2008**, *4*, 719–727.
- (33) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2008**, *112*, 6794–6799.
- (34) Kongsted, J.; Aidas, K.; Mikkelsen, K. V.; Sauer, S. P. A. *J. Chem. Theory Comput.* **2008**, *4*, 267–277.
- (35) Mulder, F. A. A.; Filatov, M. *Chem. Soc. Rev.* **2010**, *39*, 578–590.
- (36) Gregušová, A.; Perera, S. A.; Bartlett, R. J. *J. Chem. Theory Comput.* **2010**, *6*, 1228–1239.
- (37) Kupka, T.; Stachów, M.; Nieradka, M.; Kaminsky, J.; Pluta, T. *J. Chem. Theory Comput.* **2010**, *6*, 1580–1589.
- (38) Teale, A. M.; Lutnaes, O. B.; Helgaker, T.; Tozer, D. J.; Gauss, J. *J. Chem. Phys.* **2013**, *138*, 024111.
- (39) Kupka, T.; Stachów, M.; Kaminsky, J.; Sauer, S. P. A. *Magn. Reson. Chem.* **2013**, *51*, 482–489.
- (40) Kupka, T.; Stachów, M.; Nieradka, M.; Kaminsky, J.; Pluta, T.; Sauer, S. P. A. *Magn. Reson. Chem.* **2011**, *49*, 231–236.
- (41) Jameson, A. K.; Jameson, C. J. *Chem. Phys. Lett.* **1987**, *134*, 461–466.
- (42) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (43) CFOUR, a quantum chemical program package written by Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Jusélius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O'Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Stopkiewicz, S.; Tajti, A.; Vazquez, J.; Wang, F.; Watts, J. D. and the integral packages MOLECULE (Almlöf, J.; Taylor, P. R.), PROPS (Taylor, P. R.), ABACUS (Helgaker, T.; Jensen, H. J.; Jorgensen, P.; Olsen, J.), and ECP routines by Mitin, A. V.; van Wullen, C. For the current version, see <http://www.cfour.de>.
- (44) Development version of Q-Chem. www.q-chem.com.
- (45) Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 134114.
- (46) Library of exchange and correlation functionals for DFT (libxc). www.tddft.org/programs/octopus/wiki/index.php/Libxc.
- (47) Adamo, C.; Scuseria, G. E.; Barone, V. *J. Chem. Phys.* **1999**, *111*, 2889–2899.
- (48) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664–675.
- (49) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. *J. Chem. Phys.* **2001**, *115*, 9233–9242.
- (50) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.
- (51) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.
- (52) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (53) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (54) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (55) Keal, T. W.; Tozer, D. J. *J. Chem. Phys.* **2003**, *119*, 3015–3024.
- (56) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664.
- (57) Hehre, W. J.; Ditchfield, R.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1970**, *52*, 2769–2773.
- (58) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.
- (59) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797–2803.
- (60) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (61) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (62) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (63) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- (64) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (65) Feller, D. *J. Comput. Chem.* **1996**, *17*, 1571–1586.
- (66) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorathi, V.; Chase, J.; Li, J.; Windus, T. L. *J. Chem. Inf. Model.* **2007**, *47*, 1045–1052.
- (67) Maurer, M.; Ochsenfeld, C. In preparation.
- (68) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.
- (69) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793–9802.
- (70) Fink, R. F. *J. Chem. Phys.* **2010**, *133*, 174113.
- (71) Grimme, S.; Goerigk, L.; Fink, R. F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 886–906.
- (72) Ochsenfeld, C.; Koziol, F.; Brown, S. P.; Schaller, T.; Seelbach, U. P.; Klärner, F.-G. *Solid State Nucl. Magn. Reson.* **2002**, *22*, 128–153.
- (73) Zienau, J.; Kussmann, J.; Ochsenfeld, C. *Mol. Phys.* **2010**, *108*, 333–342.
- (74) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254–7260.
- (75) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, *77*, 6059–6065.

NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on January 23, 2014. Tables 36 and 37 in the Supporting Information have been modified. The correct version was published on January 31, 2014.

5.3 Paper III: "Spin component-scaled second-order Møller-Plesset perturbation theory for calculating NMR shieldings", M. Maurer and C. Ochsenfeld, *J. Chem. Theory Comput.*, 11, 37 (2015)

Reproduced with permission from *J. Chem. Theory Comput.*, 11, 37 (2015).
Copyright 2015 American Chemical Society.

Spin Component-Scaled Second-Order Møller–Plesset Perturbation Theory for Calculating NMR Shieldings

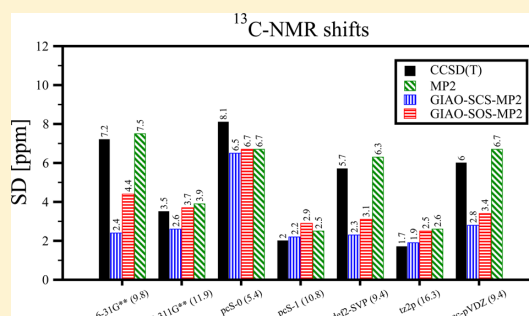
Marina Maurer and Christian Ochsenfeld*

Chair of Theoretical Chemistry, Department of Chemistry, University of Munich (LMU), Butenandtstrasse 7, D-81377 Munich, Germany

Center for Integrated Protein Science (CIPSM) at the Department of Chemistry, University of Munich (LMU), Butenandtstrasse 5–13, D-81377 Munich, Germany

Supporting Information

ABSTRACT: Spin component-scaled and scaled opposite-spin second-order Møller–Plesset perturbation approaches (SCS-MP2 and SOS-MP2) are introduced for calculating NMR chemical shifts in analogy to the well-established scaled approaches for MP2 energies. Gauge-including atomic orbitals (GIAO) are employed throughout this work. The GIAO-SCS-MP2 and GIAO-SOS-MP2 methods typically show superior performance to nonscaled MP2 and are closer to the coupled-cluster singles doubles perturbative triples (CCSD(T))/cc-pVQZ reference values. In addition, the pragmatic use of mixed basis sets for the Hartree–Fock and the correlated part of NMR chemical shift calculations is shown to be beneficial.



1. INTRODUCTION

Quantum-chemical calculations of nuclear-magnetic resonance (NMR) data have become important tools for the often complicated interpretation of experimental NMR spectra and in this way for gaining reliable structural information. While Hartree–Fock (HF)^{1–3} or density-functional theory (DFT)^{4,5} methods are quite efficient, more accurate and reliable results for NMR shieldings can be obtained by wave function-based correlation methods like Møller–Plesset second-order perturbation theory (MP2)^{6–9} or coupled-cluster (CC) theory.^{10–12} The importance of using reliable quantum-chemical methods cannot be overemphasized, since fully empirically parametrized approaches typically lack sensitivity to structural changes.^{13,14}

In order to improve the applicability of quantum-chemical NMR calculations to large molecular systems, linear- or even sublinear-scaling methods have been developed at HF and DFT levels, so that nowadays NMR spectra for molecules with more than 1000 atoms can be studied on simple workstation computers.^{15–17} In contrast, the formally steep $O(M^5)$ or at least $O(M^6)$ scaling with molecular size (M) of MP2 and CC methods, respectively, is still the limiting factor with respect to the size of the molecules which can be treated. Nevertheless, in recent years large efforts have been made for reducing the $O(M^5)$ scaling for calculating NMR shieldings at the MP2 level: For improving efficiency Gauss and Werner¹⁸ introduced a local-correlation approximation for molecular-orbital (MO) based MP2 shieldings that was recently employed by Loibl and Schütz¹⁹ for an efficient implementation and combined with a density fitting scheme. A different pathway was chosen by us in

developing a linear- or even sublinear-scaling formulation for MP2-NMR shieldings.²⁰ Here, we employ the atomic orbital (AO) basis and developed a pilot implementation that also provides evidence for the possibility of sublinear-scaling of MP2 theory for nuclei-selected NMR shieldings.²⁰ Motivated by our recent AO-based SOS-MP2 energy calculation for a DNA repair system with 2 025 atoms and 20 371 basis functions,²¹ we are currently working on an efficient implementation which will allow us to obtain NMR properties of similarly large systems.

These developments show the increasing possibilities for performing MP2 calculations on large molecules, so that the question arises if it is possible to improve the accuracy of MP2 NMR chemical-shift predictions by employing spin-component scaling similar to the scaling schemes introduced for MP2 energies by Grimme²² and also employed by others^{23–25} in scaling the same and opposite spin components of the MP2 energy differently. The original spin component scaled (SCS) MP2 energy method of Grimme, where the opposite spin component is scaled by a factor of 1.20 and the same spin component by a factor of 0.33 shows great benefits for reaction energies, barrier heights, geometries, and harmonic vibrational frequencies.²² An alternative was introduced by Head-Gordon and co-workers²⁴ as a simplification of the SCS-MP2 method with the so-called scaled opposite spin (SOS) MP2 method. Here, the same spin part is entirely neglected and only the opposite spin part of the MP2 energy scaled by a factor of 1.30.²⁴

Received: March 24, 2014

Published: December 22, 2014

The gain of accuracy of SOS-MP2 for reaction and atomization energies is similar to that of SCS-MP2, but the great advantage is the huge efficiency improvement. By combining the MP2 method with the resolution of identity approximation²⁴ and calculating only the opposite spin part of the MP2 energy, the scaling behavior of the computational cost reduces from $O(M^5)$ to $O(M^4)$. While the different scaling of the antiparallel and parallel spin component was also applied to the CIS(D)^{26,27} and CC2²⁸ methods for excited states, no studies have been published for calculating MP2 NMR shieldings so far.

In our present work, we introduce spin component-scaled MP2 schemes for calculating NMR chemical shifts in analogy to the well-established SCS- and SOS-approximations for MP2 energies.^{22–25} We always employ the gauge-including AO (GIAO)^{1,29,30} formulation that has proven to be most efficient for the calculation of NMR shieldings.^{2,31} In analogy to approximations for MP2 energies, we denote our schemes for NMR calculations by GIAO-SOS-MP2 and GIAO-SCS-MP2, respectively. We tested our new methods for MP2 based carbon, phosphorus, nitrogen, oxygen, and fluorine NMR shifts. To evaluate the accuracy of the carbon NMR shifts, we employ our recently introduced benchmark set.³² For the other nuclei, new benchmark sets are presented. After introducing the equations for optimizing scaling factors for NMR shieldings, we present first results for carbon NMR shifts employing a triple- ζ basis set (tz2p^{33,34}). Besides triple- ζ results, we also show the benefits of the scaled MP2 NMR method for relative carbon shifts calculated with medium-sized and also smaller basis sets. Furthermore, we present another pragmatic approach for increasing the accuracy of, e.g., carbon NMR chemical shifts by combining large basis HF results with the electron-correlation part described at MP2 or CCSD(T) levels using smaller basis sets. In the last chapter, the gain of accuracy is presented for phosphorus, nitrogen, oxygen, and fluorine MP2 NMR shifts.

2. OPPOSITE AND SAME SPIN TERMS OF MP2 NMR SHIELDINGS

Similar to ground state energies, MP2 based NMR shieldings can be separated into the HF contribution, and into the opposite spin (OS) and same spin (SS) terms of the perturbative second-order correction:

$$\sigma_{\text{SCS-MP2}} = \sigma_{\text{HF(basis set of SCS-MP2)}} + c_{\text{OS}} \sigma_{\text{OS}} + c_{\text{SS}} \sigma_{\text{SS}} \quad (1)$$

To obtain the scaling factors c_{OS} and c_{SS} , we fit the (absolute) MP2 NMR shieldings in a least-squares procedure to CCSD(T) results. In addition, we correct for systematic deviations of the (absolute) shieldings and therefore add a constant offset MSD_{opt} :

$$\sigma_{\text{SCS-MP2}} + \text{MSD}_{\text{opt}} \stackrel{!}{=} \sigma_{\text{CCSD(T)}} \quad (2)$$

By optimizing the absolute shieldings with the systematic deviation MSD_{opt} instead of relative shifts, the resulting scaling factors are independent of the chosen reference like, e.g., TMS. Since relative shifts are the difference of the scaled, e.g., TMS value and a scaled absolute shielding of another molecule, the MSD_{opt} value vanishes for relative shifts.

Furthermore, we also split the CCSD(T) based shielding tensor into its HF and correlation contributions

$$\sigma_{\text{CCSD(T)}} = \sigma_{\text{HF(basis set of CCSD(T))}} + \sigma_{\text{CCSD(T)-corr}} \quad (3)$$

and insert eqs 1 and 3 into eq 2

$$c_{\text{OS}} \sigma_{\text{OS}} + c_{\text{SS}} \sigma_{\text{SS}} + \text{MSD}_{\text{opt}} \stackrel{!}{=} \sigma_{\text{CCSD(T)-corr}} + \underbrace{\sigma_{\text{HF(basis set of CCSD(T))}} - \sigma_{\text{HF(basis set of SCS-MP2)}}}_{\Delta\sigma_{\text{HF}}} \quad (4)$$

Normally, the HF contributions in the CCSD(T) shieldings $\sigma_{\text{HF(basis set of CCSD(T))}}$ and in the MP2 shieldings $\sigma_{\text{HF(basis set of SCS-MP2)}}$ are the same, if the same basis set is used. Here, however, we use different basis sets for the two methods. Therefore, a deviation occurs because of different HF-based NMR shieldings included in the correlation methods, which plays an additional role in our fitting procedure. We abbreviate this HF difference by $\Delta\sigma_{\text{HF}}$:

$$c_{\text{OS}} \sigma_{\text{OS}} + c_{\text{SS}} \sigma_{\text{SS}} + \text{MSD}_{\text{opt}} \stackrel{!}{=} \sigma_{\text{CCSD(T)-corr}} + \Delta\sigma_{\text{HF}} \quad (5)$$

Since we aim for scaling factors for the entire chemical shift, we have included the HF difference throughout the fitting procedure. We have found this empirical approach to be pragmatic and useful, as illustrated below. An alternative approach would be to scale only the correlation parts. We have tested this variant for carbon NMR shifts and observe, as expected, a fit of a lower quality. The corresponding data and discussion can be found in section 4.2.

3. METHODOLOGICAL ASPECTS

To optimize the scaling factors for our GIAO-SCS-MP2 and GIAO-SOS-MP2 methods, for carbon NMR shifts we apply our benchmark set (see Figure 1) introduced by Flaig et al.,³² which

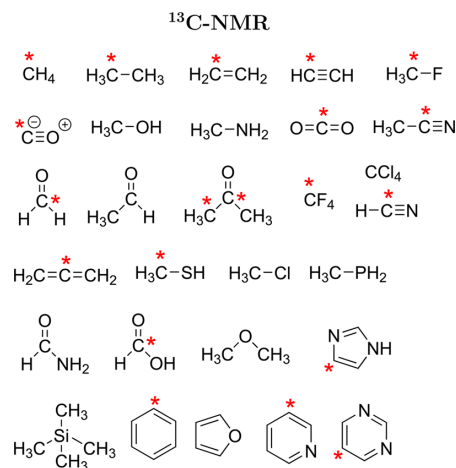


Figure 1. Molecular benchmark set of our study in ref 32. The carbon nuclei of the fitting set are labeled with a red star.

comprises various organic molecules with carbon NMR shifts spanning a broad range of the NMR scale. For the other nuclei, new benchmark sets are presented in Figures 3–6. The structures of these benchmark sets were optimized at the CCSD(T)/cc-pVTZ³⁵ level by using the program package CFOUR.³⁶ All geometries are available via the web site <http://www.cup.uni-muenchen.de/pc/ochsenfeld/download.html> and also in the Supporting Information. The reference data at the CCSD(T)/cc-pVQZ level were calculated with the program package CFOUR.³⁶

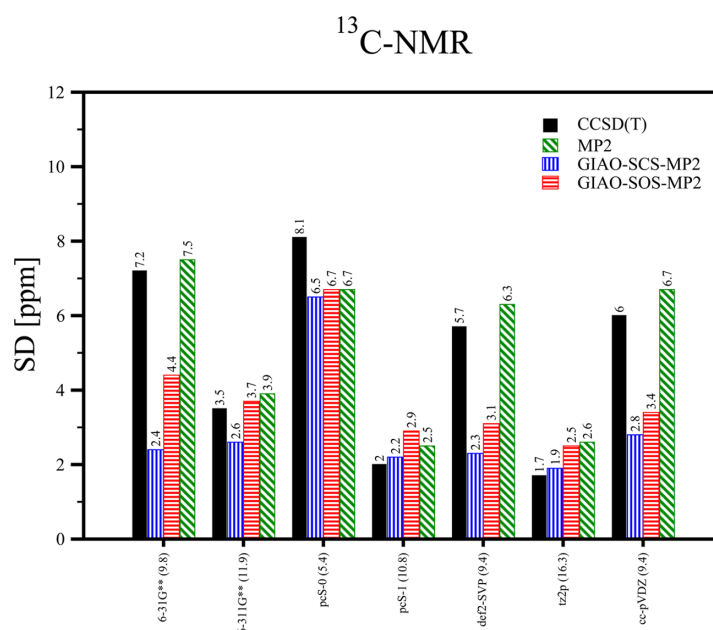


Figure 2. Standard deviation (SD) of ^{13}C chemical shifts for the GIAO-SCS-MP2 and GIAO-SOS-MP2 methods in comparison to nonscaled MP2 and CCSD(T) results (see also the work of Flaig et al.³²) are listed for various basis sets. As a reference the CCSD(T)/cc-pVQZ data is employed. Data for the total benchmark set are shown, while the fitting coefficients for GIAO-SCS-MP2 and GIAO-SOS-MP2 have been obtained for the 20-carbon shieldings fitting set. As a measure for the different basis set sizes, the average number of basis functions per atom of the molecular benchmark set is listed within parentheses.

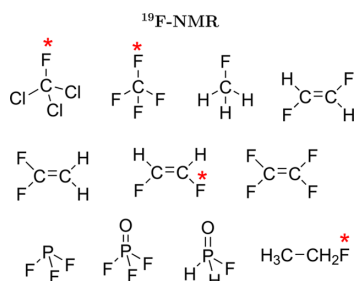


Figure 3. Molecular benchmark set for fluorine NMR shifts. The fluorine nuclei of the fitting set are labeled with a red star.

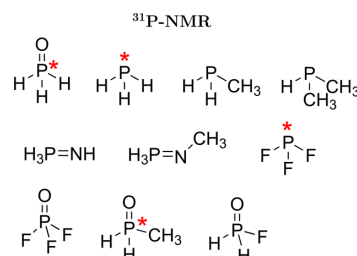


Figure 5. Molecular benchmark set for phosphorus NMR shifts. The phosphorus nuclei of the fitting set are labeled with a red star.

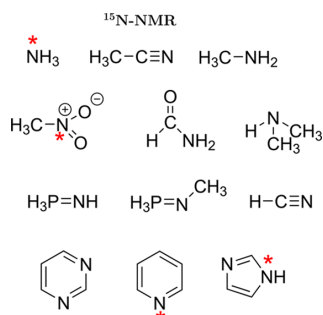


Figure 4. Molecular benchmark set for nitrogen NMR shifts. The nitrogen nuclei of the fitting set are labeled with a red star.

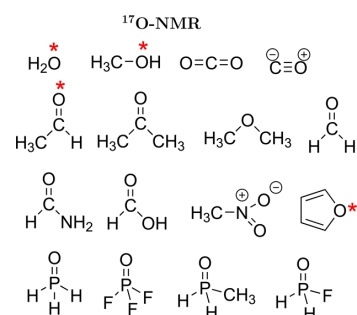


Figure 6. Molecular benchmark set for oxygen NMR shifts. The oxygen nuclei of the fitting set are labeled with a red star.

Table 1. Comparison of SD, MSD, MAD, and MAXD for GIAO-SCS-MP2 vs Nonscaled MP2 (the Difference Is Denoted by Δ) Using the tz2p Basis^a

No. of ¹³ C shieldings	Fitting Set (20)			Test Set (20)			Total Set (40)			Factors		
	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ	ϵ_{OS}	ϵ_{SS}	MSD _{opt}
SD	2.2	3.1	-0.9	1.4	2.0	-0.6	1.9	2.6	-0.7	0.856	0.632	-1.457
MSD	0.6	-0.1	0.7	-0.4	-0.8	0.4	0.1	-0.5	0.6			
MAD	1.5	2.2	-0.7	1.0	1.6	-0.6	1.3	1.9	-0.6			
MaxD	6.5	5.8	0.7	4.5	4.8	-0.3	6.5	5.8	0.7			
No. of ¹³ C shieldings	(10)			(30)			(40)			ϵ_{OS}	ϵ_{SS}	MSD _{opt}
	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ			
SD	1.8	3.0	-1.2	2.1	2.5	-0.4	2.0	2.6	-0.6	0.724	0.927	-1.271
MSD	1.0	-0.6	1.6	1.1	-0.4	1.5	1.1	-0.5	1.6			
MAD	1.4	2.3	-0.9	1.4	1.8	-0.4	1.4	1.9	-0.5			
MaxD	4.1	4.9	-0.8	8.6	5.8	2.8	8.6	5.8	2.8			
No. of ¹³ C shieldings	(5)			(35)			(40)			ϵ_{OS}	ϵ_{SS}	MSD _{opt}
	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ	GIAO-SCS-MP2	MP2	Δ			
SD	0.8	2.3	-1.5	2.0	2.6	-0.6	1.9	2.6	-0.7	0.886	0.649	-2.353
MSD	-0.2	-0.1	-0.1	0.0	-0.5	0.5	0.0	-0.5	0.5			
MAD	0.6	1.7	-1.1	1.4	1.9	-0.5	1.3	1.9	-0.6			
MaxD	1.2	3.4	-2.2	6.1	5.8	0.3	6.1	5.8	0.3			

^aData are shown for different fitting sets, test sets, and the total benchmark set (CCSD(T)/cc-pVQZ is employed as a reference). While the total benchmark set consists of 40 ¹³C shieldings, the different fitting sets vary in the number of ¹³C shieldings and the corresponding test sets comprise the remaining shieldings. In addition, the values for ϵ_{OS} , ϵ_{SS} , and MSD_{opt} are listed for the different fitting sets.

We follow the error criteria defined in previous work,³² such as the mean signed deviation (MSD), the mean absolute deviation (MAD), the standard deviation (SD), and the maximum absolute deviation (MaxD):

$$MSD = \frac{1}{N} \sum_A^N \overbrace{\delta_A - \delta_A^{CCSD(T)/cc-pVQZ}}^{\Delta\delta_A}$$

$$MAD = \frac{1}{N} \sum_A^N |\Delta\delta_A|$$

$$SD = \sqrt{\frac{1}{N-1} \sum_A^N (MSD - \Delta\delta_A)^2}$$

Here, N is the number of nuclei and δ_A the relative shielding of a nucleus A with respect to the reference nucleus such as for example TMS: $\delta_A = \sigma_{TMS} - \sigma_A$.

The opposite spin and the same spin terms of the MP2 correlation part of all nuclei were calculated with our AO-MP2 NMR method,²⁰ which is implemented in a development version of the program package Q-Chem.³⁷ These contributions were converged to a maximum deviation of 0.1 ppm with respect to MO-based MP2 shifts.

Furthermore, we employ the following basis sets: 6-31G** (6d),^{38,39} 6-311G**,⁴⁰ pcS-0, pcS-1, pcS-2,⁴¹ def2-SVP,⁴² tz2p, qz2p,^{33,34} cc-pVDZ, cc-pVTZ, and cc-pVQZ.³⁵ In all calculations, all electrons were correlated.

4. CARBON NMR SHIFTS

4.1. Accuracy Improvements of Carbon SCS-MP2 NMR Shifts for a Triple- ζ Basis Set. In the following, we discuss the influences of scaling the opposite and same spin part

of carbon MP2 NMR shieldings using the triple- ζ basis set tz2p.^{33,34} We optimized the scaling parameters for a fitting set with respect to CCSD(T)/cc-pVQZ results. The accuracy was studied for a test set with CCSD(T)/cc-pVQZ reference data. The CCSD(T)/cc-pVQZ data is the currently most reliable theoretical NMR data for the present benchmark set³² and is independent of influences from structure and vibrational effects in comparison to experimentally determined reference data.

The total benchmark set consists of 40 different carbon shieldings³² that we split into two sets each comprising 20 shieldings: one fitting set used for optimizing the scaling factors of GIAO-SCS-MP2/tz2p for approximating the CCSD(T)/cc-pVQZ values, while the other half (the test set) is used for benchmarking the obtained scaling factors. The carbon shieldings in the fitting set are labeled in Figure 1 and cover a broad range of the NMR scale for carbon nuclei (0–200 ppm) as shown in the Supporting Information.

The data shown in Table 1 for the fitting set, the test set, and the total set indicate, that the SD and the MAD are significantly reduced in comparison to nonscaled MP2. For example, the SD of the total set reduces from 2.6 to 1.9 ppm and the MAD from 1.9 to 1.3 ppm. The MSD and MaxD values are only slightly worse for some cases as, e.g., the MaxD of the total set increases from 5.8 to 6.5 ppm.

We have also investigated the influence of the number of ¹³C shieldings in the fitting set. For this purpose, instead of 20, only 10 or 5 carbon shieldings were included in the fitting set. The remaining carbon shieldings are then employed as the test set (30 or 35 carbon shieldings, respectively). As shown in Table 1, the obtained scaling factors are quite similar. Also the SD and MAD values in comparison to nonscaled MP2 are reduced.

4.2. Scaling Factors for Various Basis Sets. While the triple- ζ basis sets yield quite accurate data, the size of the basis

set plays a central role in the computation time and limits the size of the treatable systems. Therefore, the gain of accuracy by using our GIAO-SCS-MP2 and GIAO-SOS-MP2 methods for small and medium-sized basis sets is shown in Figure 2. Here, we employ the 20 carbon shieldings of the fitting set in optimizing the scaling factors to the CCSD(T)/cc-pVQZ data (the parameters c_{OS} , c_{SS} , and MSD_{opt} obtained by the fitting procedure are shown in Table 2).

Table 2. Parameters c_{OS} , c_{SS} , and MSD_{opt} for Carbon NMR Shifts As Obtained by Employing the 20-Carbon Fitting Set (Optimized Towards CCSD(T)/cc-pVQZ Results)

basis set	\bar{N}_{bas} ^a	GIAO-SCS-MP2			GIAO-SOS-MP2	
		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**	9.8	-0.029	2.193	-12.765	0.452	-13.734
6-311G**	11.9	0.647	1.236	-3.175	0.927	-4.471
pcS-0	5.4	1.219	0.689	-8.938	1.400	-9.709
pcS-1	10.8	0.885	0.886	-1.120	1.093	-2.219
def2-SVP	9.4	0.207	1.700	-10.935	0.603	-11.965
tz2p	16.3	0.856	0.632	-1.457	0.998	-2.177
cc-pVDZ	9.4	0.196	1.812	-14.020	0.614	-15.056

^aAverage number of basis functions per atom (determined for the molecular benchmark set).

The SDs of GIAO-SCS-MP2 and GIAO-SOS-MP2 are for almost all basis sets smaller than those of the nonscaled MP2 method. Exceptions are GIAO-SOS-MP2/pcS-0 for which the SD stays the same and GIAO-SOS-MP2/pcS-1 for which the SD is slightly worse. The scaling of the MP2 values shows for the medium-sized basis sets 6-31G**, def2-SVP, and cc-pVDZ huge improvements in comparison to nonscaled MP2. The SD of GIAO-SCS-MP2 using the def2-SVP basis for example reduces from 6.3 down to 2.3 ppm and is much smaller than the corresponding 5.7 ppm of CCSD(T)/def2-SVP. While the poor performance of MP2 or CCSD(T) with the small def2-SVP basis is not a surprise, the GIAO-SCS-MP2 results illustrate that the scaling parameters can somewhat (pragmatically) compensate at least some basis set deficiencies. Finally, the SDs of GIAO-SCS-MP2 for larger basis sets such as pcS-1 and tz2p are close to the SDs of the corresponding results at the CCSD(T) level, showing the usefulness of the pragmatic approach (MSD, MAD, MaxD, and the carbon shifts for all basis sets are shown in the Supporting Information).

For the different basis sets, we also investigated the approach of not including the $\Delta\sigma_{HF}$ of eq 4 in the fitting procedure. Here, we employ the same 20 carbon shieldings in the fitting set as used above (the parameters c_{OS} , c_{SS} , and MSD_{opt} are given in Table 4). The results obtained are compared with nonscaled MP2 and those of the variant with $\Delta\sigma_{HF}$ included in the fitting procedure (SDs are shown in Table 3 and the other error criteria in the Supporting Information). Also for the correlation-only scaled approach we observe, for almost all basis sets, an improvement of the SDs compared with nonscaled MP2 values. Exceptions are, again, the pcS-0 and pcS-1 basis sets. However, by neglecting the $\Delta\sigma_{HF}$ term in the fitting procedure, we obtain SDs, across all basis sets, which are higher than in the corresponding fit procedure which includes the HF-difference term. We note that the difference in the quality of the fit is heavily dependent on the size of the chosen basis set, as shown in Table 3. This observation can be traced back to the large differences in the HF-shielding of the CCSD(T)/cc-pVQZ reference calculations. Overall, our data

Table 3. Standard Deviation (SD) of ^{13}C Chemical Shifts for the GIAO-SCS-MP2 and GIAO-SOS-MP2 Methods in Comparison to Nonscaled MP2 for Various Basis Sets^a

basis set	\bar{N}_{bas} ^b	MP2	GIAO-SCS-MP2 (without $\Delta\sigma_{HF}$)	GIAO-SCS-MP2 (with $\Delta\sigma_{HF}$)	GIAO-SOS-MP2 (without $\Delta\sigma_{HF}$)	GIAO-SOS-MP2 (with $\Delta\sigma_{HF}$)
6-31G**	9.8	7.5	5.6	2.4	6.4	4.4
6-311G**	11.9	3.9	3.1	2.6	3.7	3.7
pcS-0	5.4	6.7	12.4	6.5	12.5	6.7
pcS-1	10.8	2.5	2.5	2.2	2.9	2.9
def2-SVP	9.4	6.3	3.9	2.3	4.5	3.1
tz2p	16.3	2.6	2.0	1.9	2.5	2.5
cc-pVDZ	9.4	6.7	4.2	2.8	4.7	3.4

^aFor the scaled MP2 methods two different variants are shown, where the HF deviation $\Delta\sigma_{HF}$ of eq 4 is either omitted (without $\Delta\sigma_{HF}$) or included (with $\Delta\sigma_{HF}$) in the fitting procedure. As a reference, the CCSD(T)/cc-pVQZ data is employed. ^bAverage number of basis functions per atom (determined for the molecular benchmark set).

Table 4. Parameters c_{OS} , c_{SS} , and MSD_{opt} for Carbon NMR Shifts as Obtained by Employing the 20-Carbon Fitting Set (Optimized Towards CCSD(T)/cc-pVQZ Results)^a

basis set	\bar{N}_{bas} ^b	GIAO-SCS-MP2 (without $\Delta\sigma_{HF}$)			GIAO-SOS-MP2 (without $\Delta\sigma_{HF}$)	
		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**	9.8	0.709	1.169	-1.367	0.966	-1.883
6-311G**	11.9	0.828	0.546	-0.167	0.952	-0.740
pcS-0	5.4	0.469	0.410	-1.688	0.577	-2.146
pcS-1	10.8	0.883	0.467	0.110	0.992	-0.468
def2-SVP	9.4	0.683	1.034	-0.892	0.924	-1.518
tz2p	16.3	0.949	0.395	0.143	1.038	-0.307
cc-pVDZ	9.4	0.656	1.063	-1.604	0.900	-2.211

^aHere, the HF deviation $\Delta\sigma_{HF}$ of eq 4 is omitted (without $\Delta\sigma_{HF}$) in the fitting procedure. ^bAverage number of basis functions per atom (determined for the molecular benchmark set).

strongly motivate the use of scaling factors, which are obtained from the fitting procedure, where the $\Delta\sigma_{HF}$ term is included.

4.3. Scaling Influences for Applying Different Basis Sets in the HF and in the Correlation Terms of Carbon NMR Shifts. In the previous sections, we employed the standard approach of using the same basis sets for the HF term σ_{HF} and the correlation parts of MP2 and CCSD(T) NMR shifts. Here, we propose a pragmatic alternative and less common approach to employ different basis sets for the two terms. Since HF-based NMR calculations are much faster than conventional correlation methods, larger basis sets are affordable for the HF term. As an example, an HF, MP2, or CCSD(T) calculation (with the program package CFOUR³⁶ on a single core of an Intel Xeon E5-2620 workstation using 128 GB RAM) for the molecule pyrimidine (from our benchmark set) and the pcS-1 basis set requires roughly 1, 3, or 116 min, respectively. With the larger basis set pcS-2, the calculation takes 18, 57, or 2208 min, respectively. These time differences become even larger when increasing the size of the molecule because of the different scaling behavior of these methods. As a consequence, it is typically affordable to perform an additional calculation of HF NMR shieldings with a larger basis set in contrast to the costly account of correlation effects with a larger basis set.

Table 5. Standard Deviation (SD) of ^{13}C Shifts Based on CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and Nonscaled MP2 with Respect to the CCSD(T)/cc-pVQZ Reference Data in Employing Different Basis Sets for the HF Part and the Correlation Part (corr. part)^a

basis set		CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
corr. part (\bar{N}_{bas}^b)	HF part (\bar{N}_{bas}^b)				
6-31G** (9.8)	6-31G** (9.8)	7.2	2.4	4.4	7.5
6-31G** (9.8)	qz2p (18.7)	2.2	1.7	2.6	2.6
6-31G** (9.8)	pcS-2 (23.2)	1.4	1.7	2.3	2.0
6-31G** (9.8)	cc-pVQZ (41.8)	2.2	1.5	2.5	2.7
pcS-1 (10.8)	pcS-1 (10.8)	2.1	2.2	2.9	2.5
pcS-1 (10.8)	qz2p (18.7)	1.2	1.7	2.1	2.4
pcS-1 (10.8)	pcS-2 (23.2)	0.9	2.1	2.1	2.5
pcS-1 (10.8)	cc-pVQZ (41.8)	1.4	1.6	2.0	2.6
tz2p (16.3)	tz2p (16.3)	1.7	1.9	2.5	2.6
tz2p (16.3)	qz2p (18.7)	0.8	1.6	1.9	2.1
tz2p (16.3)	pcS-2 (23.2)	1.1	1.9	2.0	2.5
tz2p (16.3)	cc-pVQZ (41.8)	0.8	1.5	1.8	2.2

^aThe new parameters for the GIAO-SCS-MP2 and GIAO-SOS-MP2 methods were optimized towards CCSD(T)/cc-pVQZ results of the 20-carbon fitting set. ^bAverage number of basis functions per atom (determined for the molecular benchmark set).

Therefore, we tested the pragmatic approach of calculating the HF part with a larger basis and the correlation part with a smaller basis set. The corresponding results shown in Table 5 illustrate the gain of accuracy for carbon NMR shifts based on the CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, or nonscaled MP2 method vs the standard approach of employing the same basis for both parts. As before, we optimized the corresponding scaling factors for our GIAO-SCS-MP2 and GIAO-SOS-MP2 methods (with a fitting set consisting of 20 carbon NMR shieldings and with CCSD(T)/cc-pVQZ as a reference). The corresponding scaling factors are shown in Table 6.

Table 6. Parameters c_{OS} , c_{SS} , and MSD_{opt} for Using Different Basis Sets for the HF Part and the Correlation Part (corr. part) of MP2 Based Carbon Shifts Optimized for the 20-Carbon Fitting Set (towards CCSD(T)/cc-pVQZ Data)

basis set		GIAO-SCS-MP2			GIAO-SOS-MP2	
corr. part	HF part	c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**	6-31G**	-0.029	2.193	-12.765	0.452	-13.734
6-31G**	qz2p	0.748	1.222	-1.041	1.016	-1.581
6-31G**	pcS-2	0.908	0.762	-0.228	1.075	-0.564
6-31G**	cc-pVQZ	0.709	1.169	-1.367	0.966	-1.883
pcS-1	pcS-1	0.885	0.886	-1.120	1.093	-2.219
pcS-1	qz2p	0.929	0.501	0.508	1.047	-0.114
pcS-1	pcS-2	1.081	0.058	1.181	1.094	1.109
pcS-1	cc-pVQZ	0.883	0.467	0.110	0.992	-0.468
tz2p	tz2p	0.856	0.632	-1.457	0.998	-2.177
tz2p	qz2p	1.001	0.418	0.532	1.095	0.056
tz2p	pcS-2	1.138	0.021	1.326	1.143	1.302
tz2p	cc-pVQZ	0.949	0.395	0.143	1.038	-0.307

By replacing for GIAO-SCS-MP2 and GIAO-SOS-MP2 the HF contribution of the MP2 part $\sigma_{\text{HF(MP2)}}$ in eq 4 with a different basis set, the described HF deviation $\Delta\sigma_{\text{HF}}$ in eq 4 changes. If one employs the cc-pVQZ basis for the HF contribution of the MP2 part, the deviation is zero ($\Delta\sigma_{\text{HF}} = 0$), since the

Table 7. Parameters c_{OS} , c_{SS} , and MSD_{opt} for Fluorine, Nitrogen, Oxygen, and Phosphorus NMR Shifts As Obtained by Employing Four Shieldings in the Fitting Set (Optimized Towards CCSD(T)/cc-pVQZ Results)

^{19}F -NMR		GIAO-SCS-MP2			GIAO-SOS-MP2	
basis set		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**		0.026	1.896	-16.567	0.522	-28.920
pcS-1		0.817	1.193	-4.949	1.173	-13.256
cc-pVDZ		-0.486	1.882	-17.281	-0.170	-29.446
^{15}N NMR		GIAO-SCS-MP2			GIAO-SOS-MP2	
basis set		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**		-0.023	2.069	-8.413	0.686	-16.749
pcS-1		0.967	-0.261	-3.221	0.878	-2.015
cc-pVDZ		0.449	0.951	-12.712	0.780	-16.913
^{17}O NMR		GIAO-SCS-MP2			GIAO-SOS-MP2	
basis set		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**		0.009	0.946	-5.979	0.210	-8.926
pcS-1		1.078	0.188	-8.742	1.107	-9.216
cc-pVDZ		0.436	0.554	-12.386	0.543	-14.166
^{31}P NMR		GIAO-SCS-MP2			GIAO-SOS-MP2	
basis set		c_{OS}	c_{SS}	MSD_{opt}	c_{OS}	MSD_{opt}
6-31G**		-1.324	3.286	-19.962	2.412	-29.188
pcS-1		-2.492	2.533	15.884	0.591	9.891
cc-pVDZ		-3.947	4.807	-32.092	1.673	-52.584

cc-pVQZ basis set was used for the reference data (CCSD(T)/cc-pVQZ). This seems to be beneficial for the scaling procedure, as for most of the basis set combinations the lowest SD values can be obtained with this basis for the HF part (see Table 5).

While the cc-pVQZ basis is rather large, lower SDs can also be obtained with smaller basis sets for the HF contribution of the MP2 part. The carbon NMR shieldings based on GIAO-SCS-MP2 show a small SD of 1.7 ppm when using the qz2p basis for the HF part and 6-31G** for the correlation part (corr. part) as compared to 2.4 ppm in the standard GIAO-SCS-MP2/6-31G** approach—see Table 5. Also the combination of corr. part:tz2p and HF part:qz2p results in a low SD value of 1.6 ppm (vs 1.9 ppm). For the GIAO-SOS-MP2 approach, the combination of corr. part:pcS-1 and HF part:qz2p is a very cost-efficient way to calculate NMR shieldings and reduces the SD to 2.1 ppm (vs 2.9 ppm). An even lower SD of 1.9 ppm (vs 2.5 ppm) can be obtained by using the combination corr. part:tz2p and HF part:qz2p.

For nonscaled MP2, using the pcS-1 basis set for both parts remains a good compromise between accuracy and cost (SD of 2.5 ppm). Although a slightly lower SD can be obtained by using the combinations corr. part:6-31G**/HF part:pcS-2 (SD of 2.0 ppm) or corr. part:tz2p/HF part:qz2p (SD of 2.1 ppm), SD values under 2 ppm however can not be achieved by using a larger basis set for the HF-part.

Using different basis sets for the HF and correlation contributions is also beneficial for CCSD(T) NMR shieldings. For example, the SD reduces from 2.1 to 0.9 ppm for the combination corr. part:pcS-1/HF part:pcS-2 and from 1.7 to 0.8 ppm for corr. part:tz2p/HF part:qz2p. Furthermore, NMR shieldings based on CCSD(T) with a SD value under 1.5 ppm can be obtained even faster, if the combinations corr. part:6-31G**/HF part:pcS-2 (vs 7.2 ppm) or corr. part:pcS-1/HF part:qz2p (vs 2.1 ppm) are applied. Further data for this approach such as the MAD and MSD of the different methods are given in the Supporting Information.

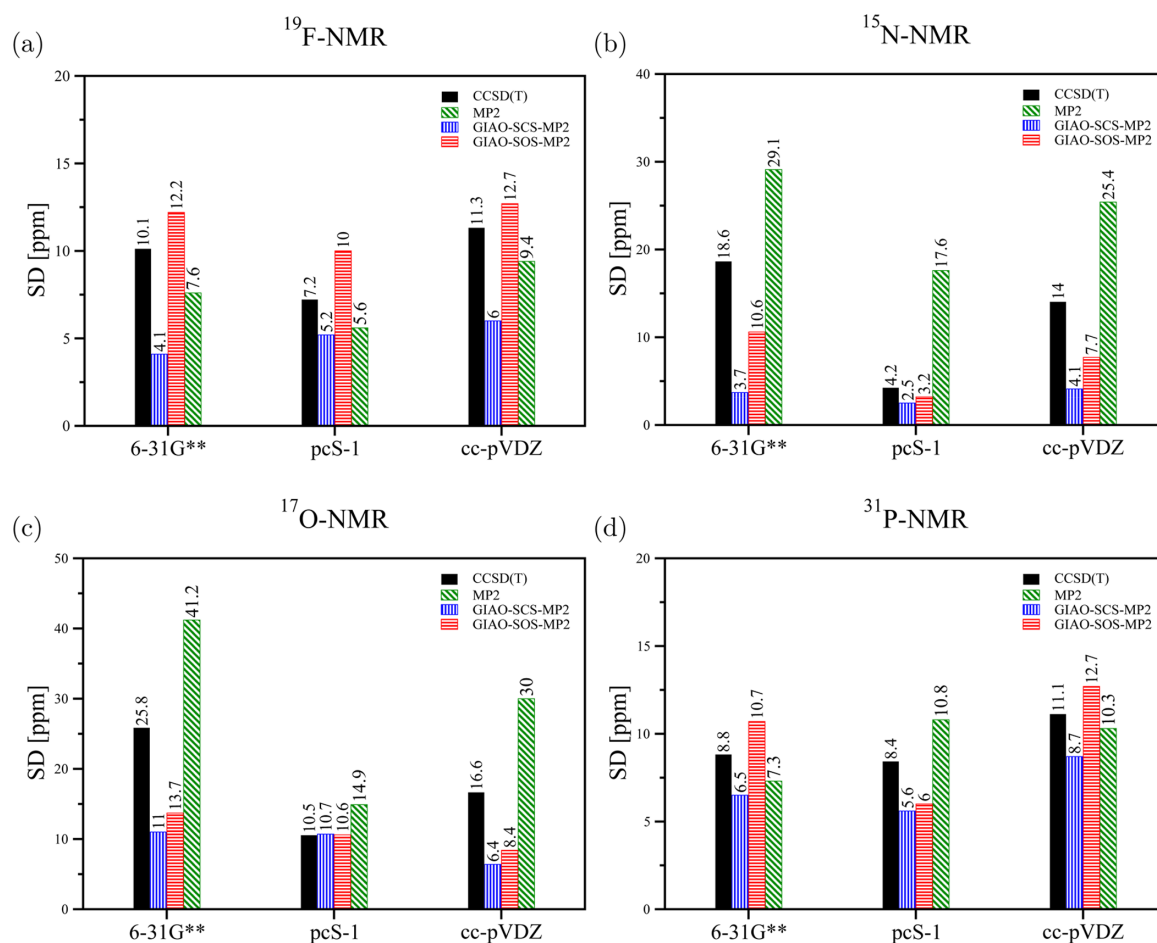


Figure 7. Standard deviation (SD) of fluorine, nitrogen, oxygen, and phosphorus chemical shifts for the GIAO-SCS-MP2 and GIAO-SOS-MP2 methods in comparison to nonscaled MP2 and CCSD(T) results are listed for different basis sets. As a reference the CCSD(T)/cc-pVQZ data is employed. Data for the total benchmark set are shown, while the fitting coefficients for GIAO-SCS-MP2 and GIAO-SOS-MP2 have been obtained with four shieldings in the fitting set.

5. PHOSPHORUS, NITROGEN, OXYGEN, AND FLUORINE NMR SHIFTS

The gain of accuracy by using our GIAO-SCS-MP2 and GIAO-SOS-MP2 methods were also investigated for phosphorus, nitrogen, oxygen, and fluorine NMR shifts. The benchmark sets of these nuclei consists of 10 phosphorus (relative to PH_3), 13 nitrogen (relative to CH_3NO_2), 18 oxygen (relative to H_2O) and 11 fluorine (relative to CFCl_3) NMR shifts. To optimize the scaling parameters with respect to CCSD(T)/cc-pVQZ results, we employ four shieldings in the fitting set in each benchmark set (the parameters c_{OS} , c_{SS} , and MSD_{opt} are shown in Table 7).

The SDs with respect to CCSD(T)/cc-pVQZ are presented for the different nuclei in Figure 7. While the SDs of GIAO-SOS-MP2 for fluorine and phosphorus nuclei are mostly larger than those of nonscaled MP2, the GIAO-SOS-MP2 shifts show huge accuracy improvements for nitrogen and oxygen. The SDs of GIAO-SCS-MP2 are for all shown basis sets smaller than those of the nonscaled MP2 method. Furthermore, for the same basis set (smaller than cc-pVQZ basis employed for

computing the CCSD(T) reference values) our GIAO-SCS-MP2 method provides NMR shift SDs which are lower than those at CCSD(T) level using the same basis set if compared to the CCSD(T)/cc-pVQZ reference value. This indicates the advantages of the empirical fitting procedure with the drawback of lost ab initio character. The SD of oxygen GIAO-SCS-MP2 shifts using the cc-pVDZ basis for example reduces from 30 to 6.4 ppm and is much smaller than the corresponding 16.6 ppm of CCSD(T)/cc-pVDZ. Furthermore, the SD of nitrogen GIAO-SCS-MP2 shifts with the basis set pcS-1 reduces from 17.6 to 2.5 ppm and the SD of phosphorus GIAO-SCS-MP2 shifts using the basis set pcS-1 from 10.8 to 5.6 ppm.

Further data such as the MAD, MSD, MaxD, and the individual shifts of the different nuclei for all basis sets are given in the Supporting Information.

6. CONCLUSION

We have introduced GIAO-SCS-MP2 and GIAO-SOS-MP2 approaches and explored their usefulness for calculating NMR chemical shifts. The results improve significantly as compared

to nonscaled MP2 and results are closer to the CCSD(T)/cc-pVQZ data used as a reference. While such scaled MP2 approaches are always empirical, they seem useful from a pragmatic point of view. Furthermore, we studied the effect of employing different basis sets for the Hartree–Fock part of MP2 and CC NMR chemical shielding calculations and show the usefulness as another pragmatic approach. The latter mixed basis approach is particularly useful if combined with scaled GIAO-MP2 approaches.

■ ASSOCIATED CONTENT

■ Supporting Information

Explicit values of the shifts determined with our GIAO-SCS-MP2 and GIAO-SOS-MP2 methods and all error criteria such as the SD, MSD, MAD, and MaxD. Furthermore, the CCSD(T)/cc-pVTZ optimized structures of the benchmark sets as well as the additional details of the approach of using different basis sets in the HF and in the correlation terms of carbon NMR shifts. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

■ AUTHOR INFORMATION

Corresponding Author

*Electronic mail: christian.ochsenfeld@uni-muenchen.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. D. Flaig and Dr. S. A. Maurer (University of Munich, LMU) for useful discussions. C.O. acknowledges financial support by the “Deutsche Forschungsgemeinschaft” (DFG) funding proposal Oc35/4-1 and the DFG cluster of excellence EXC 114 “Center for Integrated Protein Science Munich” (CIPSM).

■ REFERENCES

- (1) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789–807.
- (2) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- (3) Häser, M.; Ahlrichs, R.; Baron, H.; Weis, P.; Horn, H. *Theor. Chim. Acta* **1992**, *83*, 455–470.
- (4) Schreckenbach, G.; Ziegler, T. *J. Phys. Chem.* **1995**, *99*, 606–611.
- (5) Rauhut, G.; Puyear, S.; Wolinski, K.; Pulay, P. *J. Phys. Chem.* **1996**, *100*, 6310–6316.
- (6) Gauss, J. *Chem. Phys. Lett.* **1992**, *191*, 614–620.
- (7) Gauss, J. *J. Chem. Phys.* **1993**, *99*, 3629–3643.
- (8) Kollwitz, M.; Gauss, J. *Chem. Phys. Lett.* **1996**, *260*, 639–646.
- (9) Kollwitz, M.; Häser, M.; Gauss, J. *J. Chem. Phys.* **1998**, *108*, 8295–8301.
- (10) Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **1995**, *103*, 3561–3577.
- (11) Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **1996**, *104*, 2574–2583.
- (12) Kállay, M.; Gauss, J. *J. Chem. Phys.* **2004**, *120*, 6841–6848.
- (13) Sumowski, C. V.; Hanni, M.; Schweizer, S.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2014**, *10*, 122–133.
- (14) Christensen, A. S.; Linnet, T. E.; Borg, M.; Boomsma, W.; Lindorff-Larsen, K.; Hamelryck, T.; Jensen, J. H. *PLoS One* **2013**, *8*, e84123.
- (15) Ochsenfeld, C.; Kussmann, J.; Koziol, F. *Angew. Int. Ed.* **2004**, *43*, 4485–4489.
- (16) Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2007**, *127*, 054103.
- (17) Beer, M.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2011**, *134*, 074102.
- (18) Gauss, J.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2083.
- (19) Loibl, S.; Schütz, M. *J. Chem. Phys.* **2012**, *137*, 084107.
- (20) Maurer, M.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 174104.
- (21) Maurer, S. A.; Lambrecht, D. S.; Kussmann, J.; Ochsenfeld, C. *J. Chem. Phys.* **2013**, *138*, 014101.
- (22) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.
- (23) Grimme, S.; Goerigk, L.; Fink, R. F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 886–906.
- (24) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793–9802.
- (25) Fink, R. F. *J. Chem. Phys.* **2010**, *133*, 174113.
- (26) Grimme, S.; Izgorodina, E. I. *Chem. Phys.* **2004**, *305*, 223–230.
- (27) Rhee, Y. M.; Head-Gordon, M. *J. Phys. Chem. A* **2007**, *111*, 5314–5326.
- (28) Hellweg, A.; Grün, S. A.; Hättig, C. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4119–4127.
- (29) London, F. *J. Phys. Radium* **1937**, *8*, 397.
- (30) Hameka, H. F. *Mol. Phys.* **1958**, *1*, 203.
- (31) Gauss, J. In *Modern Methods and Algorithms of Quantum Chemistry*, ed. Grotendorst, J., Vol. 3; John von Neumann Institute for Computing: Jülich, 2000; pp 541–592.
- (32) Flaig, D.; Maurer, M.; Hanni, M.; Braunger, K.; Kick, L.; Thubauville, M.; Ochsenfeld, C. *J. Chem. Theory Comput.* **2014**, *10*, 572–578.
- (33) Auer, A. A.; Gauss, J.; Stanton, J. F. *J. Chem. Phys.* **2003**, *118*, 10407–10417.
- (34) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
- (35) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- (36) CFOUR, a quantum chemical program package written by Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G. with contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Bomble, Y. J.; Christiansen, O.; Heckert, M.; Heun, O.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Jusélius, J.; Klein, K.; Lauderdale, W. J.; Matthews, D. A.; Metzroth, T.; O’Neill, D. P.; Price, D. R.; Prochnow, E.; Ruud, K.; Schiffmann, F.; Stopkiewicz, S.; Tajti, A.; Vázquez, J.; Wang, F.; Watts, J. D. and the integral packages MOLECULE (Almlöf, J.; Taylor, P. R.), PROPS (Taylor, P. R.), ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J.), and ECP routines by Mitin, A. V. and van Wüllen, C. For the current version, see <http://www.cfour.de> (accessed November 10, 2014).
- (37) Development version of Q-Chem, www.q-chem.com (accessed November 10, 2014).
- (38) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- (39) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.
- (40) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650–654.
- (41) Jensen, F. *J. Chem. Theory Comput.* **2008**, *4*, 719–727.
- (42) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

November 27, 2014

Supporting information

**Spin component-scaled second-order Møller-Plesset
perturbation theory for calculating NMR shieldings**

Marina Maurer and Christian Ochsenfeld^{a)}

*Chair of Theoretical Chemistry, Department of Chemistry,
University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany,
and*

*Center for Integrated Protein Science (CIPSM) at the Department of Chemistry,
University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany*

1 Carbon NMR shifts

1.1 Detailed results for the basis set 6-31G**

Table 1: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-31G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 194.6 [GIAO-SCS-MP2], 190.4 [GIAO-SOS-MP2] and 207.7 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-1.7	0.3	-0.5
CH ₃ PH ₂	-1.7	-1.7	0.2	-0.4
TMS	0.0	0.0	0.0	0.0
CH ₃ CN	2.8	3.4	4.1	4.4
CH₃CH₃	8.8	9.1	9.7	9.9
CH₃SH	10.2	9.9	9.9	10.3
CH ₃ Cl	28.5	28.3	26.4	26.6
CH ₃ NH ₂	30.7	31.4	29.3	31.1
CH₃COCH₃	30.9	30.8	28.3	29.4
CH ₃ CHO	32.6	32.4	29.1	30.5
CH ₃ OH	52.0	52.4	48.1	51.3
CH ₃ OCH ₃	61.2	61.6	55.2	59.7
CHCH	71.0	71.6	66.4	62.8
CH₃F	71.1	70.4	65.0	68.7
CH ₂ CCH ₂	74.6	74.0	72.4	69.8
HCN	108.3	108.5	104.1	94.5
Furan (3, 4)	108.6	107.8	103.1	99.7
Imidazole (5)	114.6	114.6	110.5	105.7
CH₃CN	116.8	116.5	111.8	104.2
Pyrimidine (5)	122.4	123.5	114.6	117.4
CH₂CH₂	122.8	124.6	118.7	111.9
Pyridine (3, 5)	124.7	124.3	118.0	117.2
CCl ₄	126.2	122.5	106.6	115.2
CF₄	127.5	121.6	112.2	123.5
Imidazole (4)	132.5	130.2	125.3	119.9
CO₂	132.8	125.9	124.5	116.4
Benzene	133.1	130.6	127.6	121.6
Imidazole (2)	135.9	131.0	131.4	120.2
Pyridine (4)	136.6	132.4	132.9	122.8
Furan (2, 5)	139.5	137.6	132.4	127.1
Pyridine (2, 6)	153.2	148.6	147.6	138.3
Pyrimidine (4, 6)	158.8	153.6	155.1	143.0
HCONH ₂	159.6	155.4	153.1	143.1
HCOOH	161.2	156.3	153.1	144.9
Pyrimidine (2)	163.6	158.0	158.5	148.2
H₂CO	189.2	190.7	179.2	170.0
CO	189.8	188.3	188.8	165.3
CH ₃ CHO	194.4	193.1	185.2	174.8
CH₃COCH₃	201.5	198.9	193.2	181.8
CH₂CCH₂	216.5	216.1	210.5	199.4
MSD	0.0	-1.4	-4.9	-9.0
MAD	0.0	1.9	5.3	9.4
SD	0.0	2.4	4.4	7.5
MaxD	0.0	6.9	19.7	24.5

1.2 Detailed results for the basis set 6-311G**

Table 2: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-311G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 196.3 [GIAO-SCS-MP2], 193.4 [GIAO-SOS-MP2] and 199.8 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-3.5	-1.6	-3.2
CH ₃ PH ₂	-1.7	-1.5	0.1	-1.2
TMS	0.0	0.0	0.0	0.0
<u>CH₃</u> CN	2.8	2.8	3.4	3.2
CH₃CH₃	8.8	9.5	10.2	9.7
CH₃SH	10.2	10.9	11.3	11.0
CH ₃ Cl	28.5	29.6	28.6	29.1
CH ₃ NH ₂	30.7	32.1	31.0	32.3
<u>CH₃</u> CO <u>CH₃</u>	30.9	31.6	29.9	31.3
<u>CH₃</u> CHO	32.6	33.5	31.3	33.1
CH ₃ OH	52.0	52.9	50.3	53.1
CH ₃ OCH ₃	61.2	62.3	58.3	62.3
CHCH	71.0	73.7	70.4	70.8
CH₃F	71.1	70.9	67.5	70.8
<u>CH₂</u> CCH ₂	74.6	75.8	74.8	74.4
HCN	108.3	110.3	107.5	105.1
Furan (3, 4)	108.6	110.2	106.8	107.6
Imidazole (5)	114.6	117.0	113.7	114.1
<u>CH₃</u> CN	116.8	119.7	116.4	115.3
Pyrimidine (5)	122.4	127.2	120.9	126.0
CH₂CH₂	122.8	125.7	121.5	121.7
Pyridine (3, 5)	124.7	128.2	123.5	126.3
CCl ₄	126.2	122.4	112.7	121.6
CF₄	127.5	122.7	116.7	124.6
Imidazole (4)	132.5	132.7	129.0	129.4
CO₂	132.8	132.2	131.3	128.5
Benzene	133.1	134.6	131.8	131.6
Imidazole (2)	135.9	133.8	133.3	129.5
Pyridine (4)	136.6	136.0	135.5	132.3
Furan (2, 5)	139.5	141.2	137.1	137.8
Pyridine (2, 6)	153.2	151.9	150.3	148.1
Pyrimidine (4, 6)	158.8	156.6	156.5	152.4
HCONH ₂	159.6	156.4	154.4	152.0
HCOOH	161.2	158.5	156.1	154.5
Pyrimidine (2)	163.6	161.3	160.6	157.5
H₂CO	189.2	185.8	178.5	178.9
CO	189.8	189.4	189.8	180.0
<u>CH₃</u> CHO	194.4	190.1	184.8	183.8
CH₃COCH₃	201.5	196.2	192.0	190.2
CH₂CCH₂	216.5	223.7	218.8	218.0
MSD	0.0	0.1	-2.3	-2.4
MAD	0.0	2.0	2.9	3.2
SD	0.0	2.6	3.7	3.9
MaxD	0.0	7.2	13.5	11.3

1.3 Detailed results for the basis set pcS-0

Table 3: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-0). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 198.2 [GIAO-SCS-MP2], 197.4 [GIAO-SOS-MP2] and 206.7 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	0.2	0.7	0.0
CH ₃ PH ₂	-1.7	-3.5	-3.2	-3.7
TMS	0.0	0.0	0.0	0.0
<u>CH₃</u> CN	2.8	5.1	5.4	4.9
CH₃CH₃	8.8	9.6	9.8	9.5
CH₃SH	10.2	8.1	8.2	8.1
CH ₃ Cl	28.5	25.1	24.9	25.5
CH ₃ NH ₂	30.7	31.0	30.6	31.1
<u>CH₃</u> CO <u>CH₃</u>	30.9	33.6	33.0	34.0
<u>CH₃</u> CHO	32.6	34.8	33.8	35.3
CH ₃ OH	52.0	50.2	49.3	50.4
CH ₃ OCH ₃	61.2	60.0	58.5	60.4
CHCH	71.0	74.5	73.4	76.6
CH₃F	71.1	66.9	65.8	67.1
<u>CH₂</u> CCH ₂	74.6	79.9	79.7	81.1
HCN	108.3	112.6	111.7	116.4
Furan (3, 4)	108.6	111.6	110.4	113.7
Imidazole (5)	114.6	118.1	116.9	120.5
CH₃CN	116.8	124.4	123.2	127.8
Pyrimidine (5)	122.4	130.8	129.1	132.4
CH₂CH₂	122.8	124.1	122.4	127.4
Pyridine (3, 5)	124.7	130.3	129.1	132.1
CCl ₄	126.2	117.5	113.0	119.7
CF₄	127.5	145.0	142.8	144.7
Imidazole (4)	132.5	129.6	128.3	132.3
CO₂	132.8	127.6	129.1	130.7
Benzene	133.1	134.8	134.0	137.0
Imidazole (2)	135.9	131.3	131.1	134.1
Pyridine (4)	136.6	135.3	135.3	137.6
Furan (2, 5)	139.5	139.3	137.6	142.1
Pyridine (2, 6)	153.2	149.9	149.7	152.5
Pyrimidine (4, 6)	158.8	154.1	154.4	156.7
HCONH ₂	159.6	154.2	154.1	157.6
HCOOH	161.2	160.9	160.9	164.2
Pyrimidine (2)	163.6	159.3	159.7	161.7
H₂CO	189.2	177.6	175.2	183.3
CO	189.8	213.3	213.4	220.3
CH ₃ <u>CHO</u>	194.4	185.5	183.6	190.7
CH₃COCH₃	201.5	194.6	193.0	199.7
CH₂CCH₂	216.5	220.2	217.9	224.9
MSD	0.0	0.5	-0.3	2.6
MAD	0.0	4.6	4.6	4.7
SD	0.0	6.5	6.7	6.7
MaxD	0.0	23.5	23.6	30.5

1.4 Detailed results for the basis set pcS-1

Table 4: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-1). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 197.0 [GIAO-SCS-MP2], 194.9 [GIAO-SOS-MP2] and 198.6 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-2.6	-1.5	-2.8
CH ₃ PH ₂	-1.7	-1.6	-0.6	-1.7
TMS	0.0	0.0	0.0	0.0
CH₃CN	2.8	3.1	3.5	3.2
CH₃CH₃	8.8	9.5	9.9	9.5
CH₃SH	10.2	10.8	10.9	10.8
CH ₃ Cl	28.5	29.9	29.0	29.9
CH ₃ NH ₂	30.7	32.0	31.1	32.4
CH₃COCH₃	30.9	31.4	30.2	31.7
CH₃CHO	32.6	33.5	31.8	33.8
CH ₃ OH	52.0	53.3	51.3	53.9
CH ₃ OCH ₃	61.2	62.5	59.6	63.3
CHCH	71.0	73.2	70.9	72.4
CH₃F	71.1	71.9	69.4	72.6
CH₂CCH₂	74.6	76.8	76.1	76.4
HCN	108.3	109.9	107.8	108.0
Furan (3, 4)	108.6	110.2	107.8	109.8
Imidazole (5)	114.6	116.9	114.5	116.2
CH₃CN	116.8	119.9	117.5	118.6
Pyrimidine (5)	122.4	127.1	122.7	127.8
CH₂CH₂	122.8	125.8	122.7	124.7
Pyridine (3, 5)	124.7	128.3	125.0	128.4
CCl ₄	126.2	123.5	116.4	125.1
CF₄	127.5	127.0	122.7	129.1
Imidazole (4)	132.5	132.9	130.2	132.1
CO₂	132.8	134.0	133.3	132.5
Benzene	133.1	134.6	132.6	133.8
Imidazole (2)	135.9	134.4	133.9	132.7
Pyridine (4)	136.6	136.6	136.1	135.1
Furan (2, 5)	139.5	141.3	138.4	140.5
Pyridine (2, 6)	153.2	152.6	151.4	151.2
Pyrimidine (4, 6)	158.8	157.6	157.5	155.8
HCONH ₂	159.6	158.0	156.4	156.4
HCOOH	161.2	160.4	158.6	159.1
Pyrimidine (2)	163.6	162.6	162.1	161.0
H₂CO	189.2	185.9	180.5	184.2
CO	189.8	191.3	191.2	187.0
CH ₃ CHO	194.4	191.5	187.5	189.7
CH₃COCH₃	201.5	198.5	195.3	196.7
CH₂CCH₂	216.5	224.8	221.3	223.0
MSD	0.0	0.8	-1.1	0.1
MAD	0.0	1.7	1.9	2.0
SD	0.0	2.2	2.9	2.5
MaxD	0.0	8.3	9.8	6.5

1.5 Detailed results for the basis set def2-SVP

Table 5: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set def2-SVP). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 194.3 [GIAO-SCS-MP2], 191.1 [GIAO-SOS-MP2] and 205.8 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-2.1	-0.3	-1.3
CH ₃ PH ₂	-1.7	-2.5	-0.7	-1.8
TMS	0.0	0.0	0.0	0.0
<u>CH₃</u> CN	2.8	0.8	1.6	1.7
CH₃CH₃	8.8	8.2	9.0	8.8
CH₃SH	10.2	9.5	9.9	9.6
CH ₃ Cl	28.5	28.0	26.8	26.5
CH ₃ NH ₂	30.7	30.4	29.0	30.5
<u>CH₃</u> COCH₃	30.9	29.8	27.8	29.0
<u>CH₃</u> CHO	32.6	31.0	28.5	29.9
CH ₃ OH	52.0	51.0	47.7	50.7
CH ₃ OCH ₃	61.2	60.5	55.5	59.9
CHCH	71.0	72.0	68.0	65.1
CH₃F	71.1	69.3	65.1	68.6
<u>CH₂</u> CCH ₂	74.6	73.6	72.7	70.0
HCN	108.3	106.9	103.3	95.9
Furan (3, 4)	108.6	107.3	103.7	100.7
Imidazole (5)	114.6	114.7	111.3	107.5
<u>CH₃</u> CN	116.8	116.1	112.3	106.4
Pyrimidine (5)	122.4	122.8	115.9	118.6
CH₂CH₂	122.8	124.6	119.7	114.7
Pyridine (3, 5)	124.7	124.4	119.5	119.0
CCl ₄	126.2	130.9	120.5	125.6
CF₄	127.5	121.4	114.6	124.0
Imidazole (4)	132.5	130.9	127.0	122.5
CO₂	132.8	126.2	125.0	118.3
Benzene	133.1	131.4	128.9	124.0
Imidazole (2)	135.9	132.1	132.1	122.7
Pyridine (4)	136.6	133.3	133.7	124.9
Furan (2, 5)	139.5	138.3	134.0	129.9
Pyridine (2, 6)	153.2	149.7	148.8	140.9
Pyrimidine (4, 6)	158.8	154.7	155.8	145.2
HCONH ₂	159.6	154.7	153.1	144.4
HCOOH	161.2	156.0	153.7	146.6
Pyrimidine (2)	163.6	159.4	159.8	150.9
H₂CO	189.2	187.4	178.5	171.4
CO	189.8	186.9	188.1	166.9
<u>CH₃</u> CHO	194.4	192.1	186.3	177.3
CH₃COCH₃	201.5	200.0	196.1	185.7
CH₂CCH₂	216.5	219.1	214.4	205.6
MSD	0.0	-1.4	-4.1	-7.5
MAD	0.0	2.1	4.3	7.7
SD	0.0	2.3	3.1	6.3
MaxD	0.0	6.6	12.9	22.9

1.6 Detailed results for the basis set tz2p

Table 6: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set tz2p). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 196.2 [GIAO-SCS-MP2], 194.7 [GIAO-SOS-MP2] and 198.6 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-3.7	-2.7	-4.5
CH ₃ PH ₂	-1.7	-1.2	-0.3	-1.9
TMS	0.0	-0.0	0.0	0.0
<u>CH</u> ₃ CN	2.8	2.1	2.5	2.1
CH₃CH₃	8.8	8.9	9.4	8.7
CH₃SH	10.2	10.5	10.8	10.3
CH ₃ Cl	28.5	28.9	28.4	29.0
CH ₃ NH ₂	30.7	30.9	30.5	31.5
<u>CH₃CO</u> CH₃	30.9	30.5	29.7	31.1
<u>CH</u> ₃ CHO	32.6	32.0	30.9	32.8
CH ₃ OH	52.0	52.0	50.7	53.3
CH ₃ OCH ₃	61.2	60.9	58.9	62.7
CHCH	71.0	72.7	70.9	72.1
CH₃F	71.1	70.8	69.0	72.4
<u>CH</u> ₂ CCH ₂	74.6	75.5	75.1	75.0
HCN	108.3	108.6	107.0	106.2
Furan (3, 4)	108.6	109.9	108.1	109.8
Imidazole (5)	114.6	116.2	114.5	115.7
CH₃CN	116.8	118.5	116.8	116.8
Pyrimidine (5)	122.4	125.7	122.4	127.9
CH₂CH₂	122.8	124.5	122.3	123.6
Pyridine (3, 5)	124.7	127.5	125.0	128.5
CCl ₄	126.2	121.7	116.7	125.8
CF₄	127.5	126.8	123.8	130.9
Imidazole (4)	132.5	132.9	130.9	132.3
CO₂	132.8	132.5	132.0	130.3
Benzene	133.1	134.3	132.9	133.5
Imidazole (2)	135.9	134.7	134.5	132.0
Pyridine (4)	136.6	137.0	136.8	134.7
Furan (2, 5)	139.5	140.7	138.6	140.1
Pyridine (2, 6)	153.2	152.7	151.9	150.8
Pyrimidine (4, 6)	158.8	158.3	158.3	155.4
HCONH ₂	159.6	158.6	157.4	156.6
HCOOH	161.2	160.4	159.1	158.9
Pyrimidine (2)	163.6	162.7	162.4	160.4
H₂CO	189.2	185.5	181.4	184.3
CO	189.8	190.6	190.7	184.0
CH ₃ <u>CHO</u>	194.4	191.3	188.3	189.6
CH₃CO CH₃	201.5	198.5	196.1	196.6
CH₂C CH₂	216.5	223.0	220.7	220.9
MSD	0.0	0.1	-1.2	-0.5
MAD	0.0	1.3	1.7	1.9
SD	0.0	1.9	2.5	2.6
MaxD	0.0	6.5	9.5	5.8

1.7 Detailed results for the basis set cc-pVDZ

Table 7: ^{13}C NMR shift (in ppm) at GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set cc-pVDZ). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of TMS (195.6 [Ref.], 195.1 [GIAO-SCS-MP2], 191.5 [GIAO-SOS-MP2] and 210.0 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH₄	-4.3	-2.3	-0.5	-1.4
CH ₃ PH ₂	-1.7	-1.8	-0.1	-1.0
TMS	0.0	0.0	0.0	0.0
CH₃CN	2.8	3.1	3.6	4.0
CH₃CH₃	8.8	9.8	10.5	10.4
CH₃SH	10.2	10.5	10.7	10.8
CH ₃ Cl	28.5	29.6	28.0	28.2
CH ₃ NH ₂	30.7	31.5	29.7	31.5
CH₃COCH₃	30.9	31.2	28.8	30.2
CH₃CHO	32.6	33.1	30.0	31.7
CH ₃ OH	52.0	51.5	47.7	51.0
CH ₃ OCH ₃	61.2	61.5	55.7	60.4
CHCH	71.0	72.9	68.3	66.1
CH₃F	71.1	69.3	64.5	68.2
CH₂CCH₂	74.6	74.8	73.4	71.2
HCN	108.3	108.0	104.0	96.6
Furan (3, 4)	108.6	108.7	104.5	101.9
Imidazole (5)	114.6	115.8	112.0	108.5
CH₃CN	116.8	116.8	112.9	106.8
Pyrimidine (5)	122.4	125.3	117.4	120.4
CH₂CH₂	122.8	125.3	119.8	115.1
Pyridine (3, 5)	124.7	126.7	121.0	120.8
CCl ₄	126.2	135.1	123.7	129.9
CF₄	127.5	121.0	113.5	122.8
Imidazole (4)	132.5	131.1	126.7	122.6
CO₂	132.8	126.4	125.5	118.3
Benzene	133.1	133.0	130.0	125.4
Imidazole (2)	135.9	132.1	132.1	122.8
Pyridine (4)	136.6	134.2	134.3	125.8
Furan (2, 5)	139.5	139.2	134.2	130.4
Pyridine (2, 6)	153.2	150.4	149.0	141.5
Pyrimidine (4, 6)	158.8	154.9	155.8	145.5
HCONH ₂	159.6	154.5	152.6	144.0
HCOOH	161.2	156.4	153.9	146.8
Pyrimidine (2)	163.6	159.5	159.5	151.0
H₂CO	189.2	188.2	178.5	171.5
CO	189.8	188.9	189.6	168.9
CH ₃ CHO	194.4	192.4	186.0	177.1
CH₃COCH₃	201.5	199.8	195.6	185.3
CH₂CCH₂	216.5	218.2	213.0	203.9
MSD	0.0	-0.6	-3.6	-6.8
MAD	0.0	1.9	4.0	7.4
SD	0.0	2.8	3.4	6.7
MaxD	0.0	8.9	14.1	20.9

1.8 Detailed results for omitting the $\Delta \sigma_{\text{HF}}$ term in the fitting procedure

Table 8: MSD, MAD, SD, and MaxD of ^{13}C chemical shifts for the GIAO-SCS-MP2 and GIAO-SOS-MP2 methods in comparison to nonscaled MP2 are shown for various basis sets. For the scaled MP2 methods two different variants are shown, where the HF deviation $\Delta \sigma_{\text{HF}}$ is either omitted (without $\Delta \sigma_{\text{HF}}$) or included (with $\Delta \sigma_{\text{HF}}$) in the fitting procedure. As a reference the CCSD(T)/cc-pVQZ data is employed.

	MP2	GIAO-SCS-MP2 (without $\Delta \sigma_{\text{HF}}$)	GIAO-SCS-MP2 (with $\Delta \sigma_{\text{HF}}$)	GIAO-SOS-MP2 (without $\Delta \sigma_{\text{HF}}$)	GIAO-SOS-MP2 (with $\Delta \sigma_{\text{HF}}$)
6-31G**					
MSD	-9.0	-6.8	-1.4	-8.7	-4.9
MAD	9.4	7.2	1.9	9.2	5.3
SD	7.5	5.6	2.4	6.4	4.4
MaxD	24.5	16.8	6.9	19.7	19.7
6-311G**					
MSD	-2.4	-1.4	0.1	-2.5	-2.3
MAD	3.2	2.2	2.0	3.0	2.9
SD	3.9	3.1	2.6	3.7	3.7
MaxD	11.3	9.1	7.2	13.4	13.5
pcS-0					
MSD	2.6	10.5	0.5	10.0	-0.3
MAD	4.7	11.9	4.6	11.6	4.6
SD	6.7	12.4	6.5	12.5	6.7
MaxD	30.5	65.5	23.5	65.6	23.6
pcS-1					
MSD	0.1	0.5	0.8	-0.5	-1.1
MAD	2.0	1.7	1.7	1.9	1.9
SD	2.5	2.5	2.2	2.9	2.9
MaxD	6.5	8.2	8.3	10.3	9.8
def2-SVP					
MSD	-7.5	-5.0	-1.4	-6.6	-4.1
MAD	7.7	5.2	2.1	6.9	4.3
SD	6.3	3.9	2.3	4.5	3.1
MaxD	22.9	12.8	6.6	17.0	12.9
tz2p					
MSD	-0.5	-0.6	0.1	-1.4	-1.2
MAD	1.9	1.4	1.3	1.8	1.7
SD	2.6	2.0	1.9	2.5	2.5
MaxD	5.8	6.2	6.5	9.3	9.5
cc-pVDZ					
MSD	-6.8	-4.1	-0.6	-5.9	-3.6
MAD	7.4	4.6	1.9	6.3	4.0
SD	6.7	4.2	2.8	4.7	3.4
MaxD	20.9	10.7	8.9	16.4	14.1

1.9 Detailed results for using different basis sets in the HF and in the correlation terms of carbon NMR shifts

Table 9: Mean signed deviation (MSD) of ^{13}C shifts based on CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 with respect to the reference calculation CCSD(T)/cc-pVQZ for the total benchmark set are shown for different basis sets in the HF part and the correlation part (corr. part) of the carbon shifts. The fitting set for deriving the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters consists of half of the carbon shieldings in the total benchmark set (40 C shieldings) and is optimized to CCSD(T)/cc-pVQZ results.

Basis set		CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
corr. part (N_{bas}^a)	HF part (N_{bas}^a)				
6-31G** (9.8)	6-31G** (9.8)	-10.2	-1.4	-4.9	-9.0
6-31G** (9.8)	qz2p (18.7)	-1.9	1.2	-0.8	-0.7
6-31G** (9.8)	pcS-2 (23.2)	0.4	2.4	1.1	1.7
6-31G** (9.8)	cc-pVQZ (41.8)	-2.5	0.9	-1.0	-1.3
pcS-1 (10.8)	pcS-1 (10.8)	-1.7	0.8	-1.1	0.1
pcS-1 (10.8)	qz2p (18.7)	-0.5	1.4	0.4	1.3
pcS-1 (10.8)	pcS-2 (23.2)	1.8	2.5	2.4	3.6
pcS-1 (10.8)	cc-pVQZ (41.8)	-1.1	1.1	0.1	0.7
tz2p (16.3)	tz2p (16.3)	-2.2	0.1	-1.2	-0.5
tz2p (16.3)	qz2p (18.7)	-0.2	1.1	0.3	1.6
tz2p (16.3)	pcS-2 (23.2)	2.2	2.4	2.3	3.9
tz2p (16.3)	cc-pVQZ (41.8)	-0.7	0.8	0.0	1.0

^a Average number of basis functions per atom (determined for the molecular benchmark set)

Table 10: Mean absolute deviation (MAD) of ^{13}C shifts based on CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 with respect to the reference calculation CCSD(T)/cc-pVQZ for the total benchmark set are shown for different basis sets in the HF part and the correlation part (corr. part) of the carbon shifts. The fitting set for deriving the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters consists of half of the carbon shieldings in the total benchmark set (40 C shieldings) and is optimized to CCSD(T)/cc-pVQZ results.

Basis set		CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
corr. part (N_{bas}^a)	HF part (N_{bas}^a)				
6-31G** (9.8)	6-31G** (9.8)	10.6	1.9	5.3	9.4
6-31G** (9.8)	qz2p (18.7)	2.3	1.6	1.8	2.1
6-31G** (9.8)	pcS-2 (23.2)	1.1	2.5	1.9	2.1
6-31G** (9.8)	cc-pVQZ (41.8)	2.8	1.3	1.8	2.3
pcS-1 (10.8)	pcS-1 (10.8)	2.0	1.7	1.9	2.0
pcS-1 (10.8)	qz2p (18.7)	1.0	1.7	1.5	2.2
pcS-1 (10.8)	pcS-2 (23.2)	1.9	2.8	2.7	3.8
pcS-1 (10.8)	cc-pVQZ (41.8)	1.4	1.5	1.4	2.1
tz2p (16.3)	tz2p (16.3)	2.2	1.3	1.7	1.9
tz2p (16.3)	qz2p (18.7)	0.5	1.5	1.3	2.0
tz2p (16.3)	pcS-2 (23.2)	2.2	2.7	2.6	3.9
tz2p (16.3)	cc-pVQZ (41.8)	0.8	1.3	1.3	1.8

^a Average number of basis functions per atom (determined for the molecular benchmark set)

Table 11: The maximum absolute deviation (MaxD) of ^{13}C shifts based on CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 with respect to the reference calculation CCSD(T)/cc-pVQZ for the total benchmark set are shown for different basis sets in the HF part and the correlation part (corr. part) of the carbon shifts. The fitting set for deriving the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters consists of half of the carbon shieldings in the total benchmark set (40 C shieldings) and is optimized to CCSD(T)/cc-pVQZ results.

Basis set		CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
corr. part (N_{bas}^a)	HF part (N_{bas}^a)				
6-31G** (9.8)	6-31G** (9.8)	23.7	6.9	19.7	24.5
6-31G** (9.8)	qz2p (18.7)	6.8	6.9	9.3	5.7
6-31G** (9.8)	pcS-2 (23.2)	4.6	9.1	7.5	7.1
6-31G** (9.8)	cc-pVQZ (41.8)	7.1	6.7	9.2	7.1
pcS-1 (10.8)	pcS-1 (10.8)	7.1	8.3	9.8	6.5
pcS-1 (10.8)	qz2p (18.7)	3.7	7.9	6.0	7.0
pcS-1 (10.8)	pcS-2 (23.2)	3.2	9.8	9.6	11.3
pcS-1 (10.8)	cc-pVQZ (41.8)	4.6	7.6	6.1	6.9
tz2p (16.3)	tz2p (16.3)	6.7	6.5	9.5	5.8
tz2p (16.3)	qz2p (18.7)	2.8	6.7	5.5	6.9
tz2p (16.3)	pcS-2 (23.2)	4.3	8.9	8.8	11.3
tz2p (16.3)	cc-pVQZ (41.8)	3.5	6.5	5.6	6.2

^a Average number of basis functions per atom (determined for the molecular benchmark set)

2 Phosphorus, nitrogen, oxygen, and fluorine NMR shifts

2.1 Detailed results for the basis set 6-31G**

Table 12: ^{19}F NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-31G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CFCl_3 (213.0 [Ref.], 245.8 [CCSD(T)], 213.2 [GIAO-SCS-MP2], 233.8 [GIAO-SOS-MP2] and 241.5 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3F	-269.8	-232.1	-265.0	-221.7	-242.9
$\text{CH}_3\text{CH}_2\text{F}$	-214.5	-190.5	-216.5	-182.6	-199.1
trans-1,2-Difluoroethylene	-185.9	-159.3	-181.1	-152.0	-166.8
cis-1,2-Difluoroethylene	-160.6	-138.5	-155.2	-128.8	-144.7
C_2F_4	-129.7	-115.5	-131.9	-107.0	-121.6
POF_3	-82.3	-65.8	-78.9	-54.9	-72.6
1,1-Difluoroethylene	-77.3	-63.2	-67.9	-52.1	-65.4
POH_2F	-75.5	-62.3	-78.7	-57.5	-68.0
CF_4	-58.1	-49.2	-60.8	-38.5	-54.9
PF_3	-25.5	-14.3	-22.3	-9.2	-18.1
CFCl_3	0.0	0.0	0.0	0.0	0.0
MSD	0.0	17.1	1.9	25.0	11.4
MAD	0.0	17.1	3.7	25.0	11.4
SD	0.0	10.1	4.1	12.2	7.6
MaxD	0.0	37.6	9.5	48.0	26.9

Table 13: ^{31}P NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-31G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of PH_3 (632.2 [Ref.], 640.7 [CCSD(T)], 632.2 [GIAO-SCS-MP2], 617.4 [GIAO-SOS-MP2] and 646.0 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3	0.0	0.0	0.0	0.0	0.0
CH_3PH_2	80.0	64.2	71.6	64.6	66.4
PHCH_3CH_3	142.3	114.0	127.6	116.2	118.7
PH_3NCH_3	170.7	142.8	170.0	138.0	148.3
PH_3NH	174.2	145.4	167.9	143.0	150.5
PH_3O	196.2	174.6	197.8	173.5	179.7
POF_3	223.0	197.8	216.4	200.8	202.7
POH_2CH_3	226.9	200.8	225.6	200.4	206.8
POH_2F	255.0	228.2	264.0	222.5	234.5
PF_3	348.9	323.8	349.0	339.2	337.4
MSD	0.0	-22.6	-2.7	-21.9	-17.2
MAD	0.0	22.6	4.9	21.9	17.2
SD	0.0	8.8	6.5	10.7	7.3
MaxD	0.0	28.8	14.7	32.7	23.7

Table 14: ^{15}N NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-31G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CH_3NO_2 (-114.0 [Ref.], -64.4 [CCSD(T)], -115.3 [GIAO-SCS-MP2], -117.8 [GIAO-SOS-MP2] and -5.4 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3NH	-402.8	-351.1	-403.4	-391.3	-299.0
PH_3NCH_3	-388.8	-339.2	-390.0	-381.6	-286.2
NH_3	-387.1	-341.6	-388.2	-375.8	-287.3
CH_3NH_2	-372.3	-330.1	-376.9	-366.1	-275.8
NHCH_3CH_3	-361.1	-323.5	-369.3	-361.4	-268.5
HCONH_2	-282.0	-255.0	-290.2	-290.7	-198.1
Imidazole (1)	-224.9	-203.6	-222.3	-235.7	-140.4
CH_3CN	-112.4	-99.6	-120.0	-124.9	-57.5
Imidazole (3)	-104.0	-95.1	-112.0	-118.1	-43.3
HCN	-102.7	-92.4	-109.0	-116.8	-47.4
Pyrimidin	-71.3	-62.7	-74.9	-87.2	-1.5
Pyridin	-47.0	-42.9	-53.8	-62.7	9.2
CH_3NO_2	0.0	0.0	0.0	0.0	0.0
MSD	0.0	24.6	-4.1	-4.3	73.9
MAD	0.0	24.6	4.5	9.9	73.9
SD	0.0	18.6	3.7	10.6	29.1
MaxD	0.0	51.8	8.3	15.9	103.8

Table 15: ^{17}O NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set 6-31G**). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of H_2O (341.0 [Ref.], 346.4 [CCSD(T)], 339.3 [GIAO-SCS-MP2], 332.1 [GIAO-SOS-MP2] and 353.5 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3OCH_3	-11.0	-9.4	-19.3	-21.8	-8.6
PH_3O	-10.6	8.5	-3.0	-3.8	8.2
CH_3OH	-6.0	-3.5	-9.4	-9.6	-4.0
H_2O	0.0	0.0	0.0	0.0	0.0
POH_2CH_3	43.9	42.2	35.3	30.2	43.1
CO_2	107.9	95.1	106.9	102.6	92.4
POF_3	109.8	91.7	95.1	88.3	91.9
POH_2F	146.6	134.6	131.4	120.2	136.2
HCOQH	191.0	166.6	174.0	163.4	169.8
Furan	258.6	225.8	256.9	232.7	246.0
CO	392.8	353.4	394.0	384.9	352.6
HCONH_2	393.7	359.9	383.4	381.1	349.0
HCQOH	411.1	372.0	397.5	397.3	355.8
CH_3NO_2 , anti to CH	617.3	552.9	634.0	633.2	515.1
CH_3COCH_3	636.1	590.8	632.8	627.4	568.9
CH_3NO_2 , eclipsed to CH	641.1	578.3	665.1	663.6	537.2
CH_3CHO	655.1	601.3	653.2	649.1	570.5
H_2CO	718.2	652.5	724.7	718.1	608.7
MSD	0.0	-26.8	-2.4	-7.7	-36.8
MAD	0.0	29.4	8.6	12.7	39.4
SD	0.0	25.8	11.0	13.7	41.2
MaxD	0.0	65.8	23.9	27.6	109.6

2.2 Detailed results for the basis set pcS-1

Table 16: ^{19}F NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-1). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CFCl_3 (213.0 [Ref.], 225.1 [CCSD(T)], 213.0 [GIAO-SCS-MP2], 226.3 [GIAO-SOS-MP2] and 220.7 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3F	-269.8	-248.9	-265.1	-236.8	-260.4
$\text{CH}_3\text{CH}_2\text{F}$	-214.5	-202.2	-215.2	-193.3	-211.5
trans-1,2-Difluoroethylene	-185.9	-172.3	-184.1	-164.5	-180.8
cis-1,2-Difluoroethylene	-160.6	-149.6	-159.0	-141.5	-156.5
C_2F_4	-129.7	-123.3	-132.0	-116.0	-129.6
POF_3	-82.3	-78.0	-86.8	-71.3	-85.0
1,1-Difluoroethylene	-77.3	-66.4	-69.0	-59.5	-68.1
POH_2F	-75.5	-72.7	-81.3	-67.6	-78.7
CF_4	-58.1	-52.1	-59.1	-45.2	-57.5
PF_3	-25.5	-30.8	-36.4	-27.4	-35.2
CFCl_3	0.0	0.0	0.0	-0.0	0.0
MSD	0.0	7.6	-0.8	14.2	1.5
MAD	0.0	8.5	3.8	14.5	4.3
SD	0.0	7.2	5.2	10.0	5.6
MaxD	0.0	20.8	10.8	32.9	9.7

Table 17: ^{31}P NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-1). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of PH_3 (632.2 [Ref.], 623.9 [CCSD(T)], 632.0 [GIAO-SCS-MP2], 623.2 [GIAO-SOS-MP2] and 629.1 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3	0.0	0.0	0.0	0.0	0.0
CH_3PH_2	80.0	80.2	82.1	76.5	82.9
PHCH_3CH_3	142.3	141.5	142.1	133.6	147.1
PH_3NCH_3	170.7	159.9	178.2	155.8	166.2
PH_3NH	174.2	163.5	174.7	157.6	169.3
PH_3O	196.2	191.3	194.1	179.4	196.9
POF_3	223.0	231.7	217.8	212.2	236.9
POH_2CH_3	226.9	224.0	228.4	212.0	231.1
POH_2F	255.0	255.4	269.8	241.4	262.2
PF_3	348.9	366.2	348.7	344.7	381.0
MSD	0.0	-0.4	1.9	-10.4	5.6
MAD	0.0	5.7	3.4	10.4	7.5
SD	0.0	8.4	5.6	6.0	10.8
MaxD	0.0	17.3	14.8	16.8	32.1

Table 18: ^{15}N NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-1). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CH_3NO_2 (-114.0 [Ref.], -100.4 [CCSD(T)], -113.3 [GIAO-SCS-MP2], -113.0 [GIAO-SOS-MP2] and -45.9 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3NH	-402.8	-386.3	-399.8	-401.6	-339.6
PH_3NCH_3	-388.8	-373.1	-389.0	-390.2	-324.9
NH_3	-387.1	-379.2	-386.4	-388.3	-330.8
CH_3NH_2	-372.3	-362.6	-371.7	-373.4	-313.8
NHCH_3CH_3	-361.1	-351.7	-363.2	-364.5	-301.7
HCONH_2	-282.0	-272.0	-284.4	-284.3	-219.5
Imidazole (1)	-224.9	-215.4	-226.7	-224.9	-155.2
CH_3CN	-112.4	-98.4	-108.0	-107.4	-59.4
Imidazole (3)	-104.0	-96.2	-101.9	-101.0	-46.0
HCN	-102.7	-89.4	-98.4	-97.4	-47.4
Pyrimidin	-71.3	-61.3	-68.0	-66.2	-1.3
Pyridin	-47.0	-38.3	-43.1	-41.7	12.8
CH_3NO_2	0.0	0.0	0.0	0.0	0.0
MSD	0.0	10.2	1.2	1.2	56.1
MAD	0.0	10.2	2.2	2.6	56.1
SD	0.0	4.2	2.5	3.2	17.6
MaxD	0.0	16.5	4.3	5.3	70.0

Table 19: ^{17}O NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set pcS-1). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of H_2O (341.0 [Ref.], 348.0 [CCSD(T)], 342.5 [GIAO-SCS-MP2], 340.8 [GIAO-SOS-MP2] and 357.2 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3OCH_3	-11.0	-5.4	-8.0	-8.8	-3.7
PH_3O	-10.6	21.3	18.4	17.8	21.8
CH_3OH	-6.0	-1.8	-3.0	-3.1	-1.9
H_2O	0.0	0.0	0.0	0.0	0.0
POH_2CH_3	43.9	66.3	60.8	59.0	68.8
CO_2	107.9	116.1	111.8	110.9	115.3
POF_3	109.8	119.2	113.9	112.2	121.6
POH_2F	146.6	164.9	155.6	152.6	169.0
HCOQH	191.0	194.1	188.6	186.1	199.7
Furan	258.6	259.0	260.0	254.6	283.3
CO	392.8	402.6	399.9	398.0	406.5
HCONH_2	393.7	413.7	404.5	403.5	407.4
HCQOH	411.1	428.2	416.2	415.9	416.0
CH_3NO_2 , anti to CH	617.3	625.9	600.6	601.1	593.7
CH_3COCH_3	636.1	663.7	643.3	641.8	647.4
CH_3NO_2 , eclipsed to CH	641.1	651.7	623.6	623.9	616.7
CH_3CHO	655.1	682.7	656.8	655.5	659.0
H_2CO	718.2	748.0	710.9	709.1	714.2
MSD	0.0	14.2	3.2	1.9	7.7
MAD	0.0	14.2	8.1	7.6	13.5
SD	0.0	10.5	10.7	10.6	14.9
MaxD	0.0	32.0	29.0	28.4	32.4

2.3 Detailed results for the basis set cc-pVDZ

Table 20: ^{19}F NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set cc-pVDZ). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CFCl_3 (213.0 [Ref.], 242.3 [CCSD(T)], 213.3 [GIAO-SCS-MP2], 233.2 [GIAO-SOS-MP2] and 239.2 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3F	-269.8	-233.6	-266.0	-224.6	-243.7
$\text{CH}_3\text{CH}_2\text{F}$	-214.5	-190.3	-216.5	-185.1	-198.3
trans-1,2-Difluoroethylene	-185.9	-160.3	-182.7	-155.0	-168.1
cis-1,2-Difluoroethylene	-160.6	-139.7	-155.4	-130.1	-146.0
C_2F_4	-129.7	-118.8	-133.2	-109.1	-124.8
POF_3	-82.3	-71.2	-80.0	-55.0	-78.4
1,1-Difluoroethylene	-77.3	-63.0	-65.1	-50.8	-64.6
POH_2F	-75.5	-67.5	-86.2	-66.7	-73.5
CF_4	-58.1	-53.5	-59.8	-37.1	-58.8
PF_3	-25.5	-23.8	-30.2	-15.9	-28.8
CFCl_3	0.0	0.0	0.0	0.0	0.0
MSD	0.0	14.3	0.4	22.7	8.6
MAD	0.0	14.3	4.5	22.7	9.3
SD	0.0	11.3	6.0	12.7	9.4
MaxD	0.0	36.2	12.3	45.2	26.0

Table 21: ^{31}P NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set cc-pVDZ). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of PH_3 (632.2 [Ref.], 665.9 [CCSD(T)], 632.2 [GIAO-SCS-MP2], 613.2 [GIAO-SOS-MP2] and 671.5 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3	0.0	0.0	0.0	0.0	0.0
CH_3PH_2	80.0	67.8	77.7	67.1	70.2
PHCH_3CH_3	142.3	120.4	135.9	119.4	125.1
PH_3NCH_3	170.7	138.7	178.4	136.2	145.6
PH_3NH	174.2	141.9	171.9	140.1	148.4
PH_3O	196.2	166.7	196.1	163.7	173.6
POF_3	223.0	204.2	219.1	199.5	209.3
POH_2CH_3	226.9	195.4	227.0	192.7	203.2
POH_2F	255.0	225.5	279.1	218.8	233.4
PF_3	348.9	336.7	348.9	339.6	351.3
MSD	0.0	-22.0	1.7	-24.0	-15.7
MAD	0.0	22.0	4.7	24.0	16.2
SD	0.0	11.1	8.7	12.7	10.3
MaxD	0.0	32.3	24.1	36.3	25.8

Table 22: ^{15}N NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set cc-pVDZ). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of CH_3NO_2 (-114.0 [Ref.], -74.4 [CCSD(T)], -114.4 [GIAO-SCS-MP2], -115.7 [GIAO-SOS-MP2] and -12.7 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
PH_3NH	-402.8	-361.9	-396.9	-390.9	-307.8
PH_3NCH_3	-388.8	-348.7	-383.4	-379.5	-293.3
NH_3	-387.1	-358.3	-387.5	-381.1	-302.2
CH_3NH_2	-372.3	-342.4	-372.2	-366.7	-286.0
NHCH_3CH_3	-361.1	-333.1	-362.9	-359.1	-275.7
HCONH_2	-282.0	-263.7	-288.3	-288.5	-204.5
Imidazole (1)	-224.9	-208.7	-223.8	-230.3	-142.3
CH_3CN	-112.4	-104.1	-119.2	-121.9	-58.9
Imidazole (3)	-104.0	-97.9	-108.8	-112.2	-42.4
HCN	-102.7	-95.9	-108.6	-112.5	-47.6
Pyrimidin	-71.3	-64.6	-72.9	-79.2	0.8
Pyridin	-47.0	-42.4	-49.3	-54.3	13.9
CH_3NO_2	0.0	0.0	0.0	0.0	0.0
MSD	0.0	18.1	-1.4	-1.5	70.0
MAD	0.0	18.1	3.3	6.9	70.0
SD	0.0	14.0	4.1	7.7	25.4
MaxD	0.0	40.9	6.8	11.9	95.5

Table 23: ^{17}O NMR shift (in ppm) at CCSD(T), GIAO-SCS-MP2, GIAO-SOS-MP2, and nonscaled MP2 levels (basis set cc-pVDZ). MSD, MAD, SD, and MaxD are determined with respect to CCSD(T)/cc-pVQZ results (Ref.). All shifts are relative to the corresponding absolute shielding of H_2O (341.0 [Ref.], 354.9 [CCSD(T)], 342.8 [GIAO-SCS-MP2], 338.1 [GIAO-SOS-MP2] and 363.0 [MP2] ppm). The nuclei of the molecules for which absolute shieldings were used in the fitting procedure to optimize the GIAO-SCS-MP2 and GIAO-SOS-MP2 parameters are labeled boldface.

Molecule	Ref.	CCSD(T)	GIAO-SCS-MP2	GIAO-SOS-MP2	MP2
CH_3OCH_3	-11.0	-1.0	-8.4	-10.3	0.4
PH_3O	-10.6	17.3	7.6	7.1	15.7
CH_3OH	-6.0	1.7	-2.4	-2.7	1.4
H_2O	0.0	0.0	-0.0	0.0	0.0
POH_2CH_3	43.9	58.9	49.6	45.7	59.8
CO_2	107.9	103.6	109.0	106.6	100.9
POF_3	109.8	109.3	108.2	104.1	109.0
POH_2F	146.6	151.9	143.5	136.2	153.4
HCOQH	191.0	178.2	179.0	172.4	181.7
Furan	258.6	245.3	260.4	244.9	267.7
CO	392.8	371.4	393.9	388.7	371.2
HCONH_2	393.7	382.1	392.2	390.1	373.8
HCOQH	411.1	395.7	406.6	406.5	380.9
CH_3NO_2 , anti to CH	617.3	584.1	623.8	624.8	546.4
CH_3COCH_3	636.1	618.6	636.1	632.1	602.1
CH_3NO_2 , eclipsed to CH	641.1	609.8	651.4	651.9	569.1
CH_3CHO	655.1	634.9	657.1	653.7	610.5
H_2CO	718.2	694.2	725.0	719.9	659.3
MSD	0.0	-7.8	2.1	-1.3	-16.2
MAD	0.0	15.1	4.6	6.1	24.8
SD	0.0	16.6	6.4	8.4	30.0
MaxD	0.0	33.1	18.2	18.6	72.0

3 The geometries of the benchmark sets

Table 24: The coordinates (in Å) of the CCSD(T)/cc-pVTZ optimized structures used in the benchmark calculations.

CHCH			
H	0.000000	0.000000	3.323548
C	0.000000	0.000000	2.264681
C	0.000000	0.000000	1.058867
H	0.000000	0.000000	0.000000
CH ₂ CH ₂			
H	-0.920801	0.000000	1.895032
H	0.920801	0.000000	1.895032
C	0.000000	0.000000	1.332960
H	-0.920801	0.000000	-0.562071
C	0.000000	0.000000	0.000000
H	0.920801	0.000000	-0.562071
CH ₃ CH ₃			
H	-1.015004	0.000000	1.915222
H	0.507502	0.879020	1.915222
H	0.507502	-0.879019	1.915222
C	0.000000	0.000000	1.522992
H	-0.507502	0.879019	-0.392230
H	-0.507502	-0.879019	-0.392230
C	0.000000	0.000000	0.000000
H	1.015004	0.000000	-0.392230
Benzene			
H	0.000000	0.000000	-2.497610
H	2.162994	0.000000	-1.248805
C	0.000000	0.000000	-1.403792
H	-2.162994	0.000000	-1.248805
C	1.215720	0.000000	-0.701896
H	2.162994	0.000000	1.248805
C	-1.215720	0.000000	-0.701896
H	-2.162994	0.000000	1.248805
C	1.215720	0.000000	0.701896
C	-1.215720	0.000000	0.701896
C	0.000000	0.000000	1.403792
H	0.000000	0.000000	2.497610
CCl ₄			
C	0.000000	0.000000	0.000000
Cl	0.000000	-1.446410	1.022766
Cl	0.000000	1.446410	1.022766
Cl	1.446410	0.000000	-1.022766
Cl	-1.446410	0.000000	-1.022766

CF ₄			
F	0.000000	0.000000	1.316355
F	-0.620536	1.074799	-0.438785
F	-0.620536	-1.074799	-0.438785
C	0.000000	0.000000	0.000000
F	1.241071	0.000000	-0.438785
CH ₂ CCH ₂			
H	1.861493	-0.000026	0.074137
H	1.861493	0.000027	1.925863
C	1.308022	0.000000	1.000000
C	0.000000	0.000000	1.000000
H	-1.861493	0.925863	0.999974
C	-1.308022	0.000000	1.000000
H	-1.861493	-0.925863	1.000025
CH ₃ CHO			
O	0.996330	0.000000	2.182240
H	-0.996558	0.000000	1.968338
C	0.000000	0.000000	1.498621
H	-0.535470	0.876780	-0.363023
H	-0.535470	-0.876780	-0.363023
C	0.000000	0.000000	0.000000
H	1.017087	0.000000	-0.377116
CH ₃ Cl			
C	1.227632	0.000000	0.000000
Cl	-0.557102	0.000000	0.000000
H	1.570923	-0.513161	-0.888821
H	1.570923	1.026323	0.000000
H	1.570923	-0.513161	0.888821
CH ₃ CN			
N	-1.227400	2.139471	-0.878717
C	-0.684431	1.193026	-0.489740
H	1.024493	0.000000	-0.359635
H	-0.509002	-0.887237	-0.364213
C	0.000000	0.000000	0.000000
H	0.000000	0.000000	1.085783
CH ₃ COCH ₃			
H	-1.237388	0.000000	3.250316
H	-1.920322	0.876759	1.860539
H	-1.920322	-0.876759	1.860539
C	-1.354117	0.000000	2.172571
O	1.028727	0.000000	2.149464
C	0.000000	0.000000	1.508281
H	-0.529512	0.876759	-0.370905
H	-0.529512	-0.876759	-0.370905
C	0.000000	0.000000	0.000000
H	1.018997	0.000000	-0.369871

CH ₃ F			
F	0.000000	0.000000	1.378350
H	-0.513206	0.888898	-0.356755
H	-0.513206	-0.888898	-0.356755
C	0.000000	0.000000	0.000000
H	1.026411	0.000000	-0.356755
CH ₃ NH ₂			
H	0.953113	0.000000	1.799235
H	-0.410956	0.859965	1.799235
N	0.000000	0.000000	1.462391
H	0.528050	0.837588	-0.463821
H	-1.027225	0.011085	-0.356303
C	0.000000	0.000000	0.000000
H	0.452913	-0.922055	-0.356303
CH ₃ OCH ₃			
H	1.033985	0.000000	1.735913
H	-0.501256	-0.887279	1.805175
H	-0.501256	0.887279	1.805175
C	0.000000	0.000000	1.406411
O	0.000000	0.000000	0.000000
H	-1.865505	0.887279	-0.172642
H	-1.865505	-0.887279	-0.172642
C	-1.314696	0.000000	-0.499566
H	-1.255432	0.000000	-1.583165
CH ₃ OH			
H	0.913072	0.000000	1.705672
O	0.000000	0.000000	1.416457
H	0.481572	-0.887133	-0.414332
H	0.481572	0.887133	-0.414332
C	0.000000	0.000000	0.000000
H	-1.038584	0.000000	-0.315160
CH ₃ PH ₂			
C	-1.261578	0.000000	0.015592
P	0.595347	0.000000	-0.063914
H	0.817833	1.028354	0.882260
H	0.817833	-1.028354	0.882260
H	-1.631757	0.877968	-0.507458
H	-1.631757	-0.877968	-0.507459
H	-1.647733	-0.000000	1.029013
CH ₃ SH			
C	-1.230645	0.011288	0.000000
S	0.587929	-0.043995	0.000000
H	0.782250	1.279595	-0.000000
H	-1.564501	-1.021341	0.000000
H	-1.608023	0.501510	0.890190
H	-1.608023	0.501509	-0.890190

CH ₄			
H	0.000000	0.000000	1.085884
H	-0.511891	0.886621	-0.361961
H	-0.511891	-0.886621	-0.361961
C	0.000000	0.000000	0.000000
H	1.023781	0.000000	-0.361961
CO ₂			
O	0.000000	0.000000	-1.163261
C	0.000000	0.000000	0.000000
O	0.000000	0.000000	1.163261
CO			
O	0.000000	0.000000	1.132467
C	0.000000	0.000000	0.000000
Furan			
H	1.539212	0.000000	2.843142
H	-0.954720	0.000000	1.815637
C	1.248435	0.000000	1.815766
H	3.170737	0.000000	0.631713
C	0.000000	0.000000	1.340330
C	2.103305	0.000000	0.657044
O	0.000000	0.000000	0.000000
C	1.280647	0.000000	-0.395508
H	1.453069	0.000000	-1.447971
H ₂ CO			
O	1.206499	0.000000	1.000000
H	-0.580683	0.932484	1.016277
C	0.000000	0.000000	1.000000
H	-0.580683	-0.932484	0.983723
HCN			
H	0.000000	0.000000	-1.061390
C	0.000000	0.000000	0.000000
N	0.000000	0.000000	1.156836
HCONH ₂			
H	-0.340220	-1.983843	1.349805
H	-1.774691	-1.028891	1.181421
N	-0.776727	-1.094600	1.193007
H	-0.595137	0.907846	0.839922
C	0.000000	0.000000	1.000000
O	1.210759	0.000000	1.000000
HCOOH			
H	-0.167681	-1.826330	1.322031
O	-0.771229	-1.082419	1.190860
H	-0.626930	0.877038	0.845355
C	0.000000	0.000000	1.000000
O	1.199044	0.000000	1.000000

Imidazole			
H	1.691702	0.000000	-1.292070
H	3.126870	0.000000	0.790475
N	1.327772	0.000000	-0.358893
H	-0.789482	0.000000	-0.723323
C	2.054013	0.000000	0.791544
C	0.000000	0.000000	0.000000
N	1.287878	0.000000	1.857017
C	0.000000	0.000000	1.367668
H	-0.848690	0.000000	2.022934
Pyridin			
H	2.050022	0.000000	1.193993
N	0.000000	0.000000	1.315837
H	-2.050022	0.000000	1.193993
C	1.138130	0.000000	0.615031
H	2.144544	0.000000	-1.278934
C	-1.138130	0.000000	0.615031
H	-2.144544	0.000000	-1.278934
C	1.193321	0.000000	-0.774669
C	-1.193321	0.000000	-0.774669
C	0.000000	0.000000	-1.485408
H	0.000000	0.000000	-2.562993
Pyrimidin			
H	-2.053215	0.000000	1.259948
N	0.000000	0.000000	1.399515
H	2.048054	0.000000	1.238005
C	-1.134197	0.000000	0.693965
H	-2.067337	0.000000	-1.249692
C	1.124377	0.000000	0.679659
C	-1.146976	0.000000	-0.693341
N	1.239185	0.000000	-0.650473
C	0.087332	0.000000	-1.326787
H	0.161353	0.000000	-2.403946

TMS			
H	1.017146	0.000000	2.270427
H	-0.508573	-0.880874	2.270427
H	-0.508573	0.880874	2.270427
H	1.917911	1.560171	-0.277322
H	0.392192	2.441045	-0.277322
H	0.900765	1.560171	-1.715783
H	-2.310103	0.880874	-0.277322
H	-2.310103	-0.880874	-0.277322
H	-1.801530	0.000000	-1.715783
C	0.000000	0.000000	1.879309
C	0.885915	1.534449	-0.626436
C	-1.771829	0.000000	-0.626436
Si	0.000000	0.000000	0.000000
H	1.917911	-1.560171	-0.277322
H	0.900765	-1.560171	-1.715783
C	0.885915	-1.534449	-0.626436
H	0.392192	-2.441045	-0.277322

1,1-Difluoroethylene			
C	0.00000017	-0.00000004	-1.46246898
C	0.00000003	0.00000021	-0.14314343
F	1.07715007	-0.00000002	0.61262307
H	0.93402072	-0.00000065	-1.98961418
F	-1.07715022	-0.00000006	0.61262280
H	-0.93402023	0.00000030	-1.98961444
cis-1,2-Difluoroethylene			
C	0.00000003	-0.66278545	0.63864193
C	-0.00000005	0.66278544	0.63864191
H	-0.00000005	1.23744508	1.54590444
H	0.00000018	-1.23744515	1.54590443
F	0.00000002	1.38094147	-0.48539364
F	-0.00000002	-1.38094146	-0.48539365
trans-1,2-Difluoroethylene			
C	0.52322213	0.40620781	0.00000021
C	-0.52322213	-0.40620781	0.00000006
H	-0.46779886	-1.47886131	0.00000025
F	1.76402305	-0.09690585	-0.00000010
F	-1.76402305	0.09690585	-0.00000007
H	0.46779886	1.47886131	-0.00000025

C ₂ F ₄			
C	-0.66100259	-0.00000001	-0.00000002
C	0.66100257	0.00000001	0.00000022
F	1.38189201	1.09728861	-0.00000011
F	-1.38189200	1.09728863	0.00000005
F	1.38189200	-1.09728861	0.00000001
F	-1.38189200	-1.09728862	-0.00000007
CFCl ₃			
C	0.24641056	0.00000000	0.00000000
F	1.58149905	0.00000000	0.00000000
Cl	-0.31459314	-0.83757479	-1.45072210
Cl	-0.31459314	1.67514959	0.00000000
Cl	-0.31459314	-0.83757479	1.45072210
CH ₃ CH ₂ F			
C	-1.28522104	-0.25363812	0.00000000
C	-0.02768880	0.57519928	0.00000001
H	-1.32198872	-0.88546445	0.88356984
H	-1.32198853	-0.88546434	-0.88356980
H	-2.15822922	0.39661282	-0.00000012
F	1.08057958	-0.25798395	-0.00000001
H	0.03245370	1.20438527	0.88523477
H	0.03245360	1.20438526	-0.88523473
PF ₃			
P	0.00000000	-0.00000002	-0.50601650
F	0.00001375	1.36611601	0.27499213
F	1.18308434	-0.68306990	0.27499214
F	-1.18309809	-0.68304608	0.27499214
POF ₃			
P	-0.16278486	0.00000000	0.00000000
O	-1.60701340	0.00000000	0.00000000
F	0.53945063	0.68199760	-1.18125450
F	0.53945063	-1.36399521	0.00000000
F	0.53945063	0.68199760	1.18125450
POH ₂ F			
P	-0.11040556	0.38241453	0.00000000
O	-1.30249084	-0.46868910	0.00000000
F	1.26938697	-0.35752522	0.00000000
H	0.06775972	1.21262799	1.11087105
H	0.06775972	1.21262799	-1.11087105

PH ₃ NCH ₃			
P	-0.92632046	0.08616162	-0.00000003
N	0.48921292	-0.56714271	0.00000009
H	-1.35751392	0.92040048	1.06951314
H	-1.90557299	-0.90636728	0.00000025
H	-1.35751369	0.92039957	-1.06951357
C	1.69387339	0.24643324	0.00000003
H	2.55768799	-0.41468163	-0.00000017
H	1.78290793	0.88903109	0.88232324
H	1.78290835	0.88903026	-0.88232376
PH ₃ NH			
P	-0.44770928	0.00741099	0.00000001
N	1.11607815	-0.06866067	0.00000000
H	-1.17749261	0.59100276	1.06714780
H	-0.97023562	-1.28639733	-0.00000035
H	-1.17749237	0.59100323	-1.06714779
H	1.57761151	0.83062287	0.00000005
PH ₃ O			
P	0.43545289	0.00000000	-0.00000002
H	1.07945999	-0.35927069	-1.19940696
H	1.07946023	1.21835247	0.28856606
H	1.07946044	-0.85908169	0.91084112
O	-1.04729131	0.00000000	0.00000003
PH ₃			
P	-0.06829281	0.00000007	0.00000000
H	0.69961952	-1.18760404	0.00000000
H	0.69962095	0.59380092	-1.02849445
H	0.69962095	0.59380092	1.02849445
PHCH ₃ CH ₃			
P	0.00000025	0.59910148	-0.05406170
C	1.40967394	-0.60110398	0.01235824
H	0.00000030	0.93327453	1.32211462
C	-1.40967448	-0.60110356	0.01235824
H	-1.48384681	-1.10518180	-0.94899307
H	-2.33971091	-0.06324003	0.17851494
H	-1.28328912	-1.34714950	0.79301982
H	2.33971058	-0.06324106	0.17851594
H	1.48384668	-1.10518166	-0.94899334
H	1.28328792	-1.34715055	0.79301913

POH ₂ CH ₃			
P	-0.10876452	0.38770787	0.00000001
H	-0.02245933	1.28492863	-1.08415263
H	-0.02245965	1.28492818	1.08415278
C	1.52334686	-0.38131929	0.00000001
O	-1.27897619	-0.52452616	-0.00000002
H	2.29533253	0.38605816	0.00000074
H	1.62616352	-1.00326561	-0.88481429
H	1.62616299	-1.00326680	0.88481343
CH ₃ NO ₂			
C	1.39222996	0.07645002	-0.00000001
N	-0.09185901	-0.00103048	-0.00000007
H	1.73980266	-0.44077271	0.88588805
H	1.67668640	1.11871572	-0.00000293
H	1.73980298	-0.44077811	-0.88588477
O	-0.57651481	-1.12066810	0.00000003
O	-0.71246307	1.04927095	0.00000002
NH ₃			
N	0.00000001	-0.06928928	0.00000000
H	-0.93296178	0.32090983	0.00000000
H	0.46648086	0.32090986	0.80796860
H	0.46648086	0.32090986	-0.80796861
NHCH ₃ CH ₃			
N	-0.55651568	-0.08569524	0.00000000
H	-1.23077855	0.66720910	0.00000000
C	0.25700324	0.01595803	1.20073656
C	0.25700324	0.01595803	-1.20073656
H	-0.37980407	0.00313508	2.08159683
H	0.88357760	0.91675135	1.23491274
H	0.91773408	-0.84816064	1.25567613
H	-0.37980407	0.00313508	-2.08159683
H	0.91773408	-0.84816064	-1.25567613
H	0.88357760	0.91675135	-1.23491274
H ₂ O			
O	0.00000000	0.00000000	0.06619561
H	0.00000000	-0.75337123	-0.52528621
H	0.00000000	0.75337123	-0.52528621

- 5.4 Manuscript IV: "QQR-type integral estimates for calculating second-order Møller-Plesset perturbation theory-based NMR shieldings of selected nuclei with a sublinear-scaling computational effort",**
M. Maurer and C. Ochsenfeld,
(in preparation)

Comment

This paper is currently in preparation. In this work, the theory of the QQR-type estimates for AO-MP2 NMR shieldings are presented. While the efficient implementation of the equations is not yet complete, the preliminary results for larger molecules are mainly shown for the small basis set STO-3G. For the smaller molecules, the results for the basis set 6-31G* are also shown. Since the results of the two basis sets are similar for the smaller molecules, one can expect the same behavior for the larger molecules. Calculations of larger molecules with larger basis sets is part of future work.

**QQR-type integral estimates for calculating second-order
Møller-Plesset perturbation theory-based NMR shieldings of
selected nuclei with a sublinear-scaling computational effort**

Marina Maurer and Christian Ochsenfeld^{a)}

*Chair of Theoretical Chemistry, Department of Chemistry,
University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany*

and

*Center for Integrated Protein Science (CIPSM) at the Department of Chemistry,
University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany*

QQR-type estimates are applied for calculating NMR shieldings with atomic orbital-based second-order Møller-Plesset perturbation theory (AO-MP2). Thereby, the computational cost of the rate-determining steps of AO-MP2 NMR calculations can be reduced to sublinear for single nuclei, which is demonstrated for linear alkanes and amylose chains. Furthermore, the deviations between our AO-MP2 NMR shieldings and conventional molecular-orbital (MO)-MP2 NMR shieldings are determined, which can be reliably controlled by a screening threshold.

^{a)}Electronic mail: christian.ochsenfeld@uni-muenchen.de

1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most widely used experimental method for gaining either insights in the structure and dynamics of unknown compounds or to confirm the identity of reaction products. The information has therefore to be extracted from a spectrum containing the NMR shifts of all nuclei, which becomes more and more complicated with increasing size and complexity of the molecular system. Quantum chemical calculations of the NMR spectra can help with the interpretation as it provides complementary information not available from experiment. While Hartree-Fock (HF-) or density-functional theory (DFT-) based NMR shifts calculations can nowadays be performed for systems with more than 1000 atoms [1, 2], such system sizes are not yet accessible for second-order Møller-Plesset perturbation (MP2) theory based NMR shifts. Since this level of theory has often proven to provide more accurate results [3–6], the reduction of the computational effort for accessing larger molecules is highly desirable and the major goal of our present work.

To be able to handle large systems, the formally high computational cost of the MP2 method, which is increasing with the fifth power of the system size, must be reduced. In 2013, we presented the theory of the atomic-orbital (AO-) based MP2 NMR shieldings with a linear-scaling computational effort for all and a sublinear-scaling for computing single nuclei NMR shieldings. The advantage of the latter method is that in large molecules only the nuclei of interest can be specifically calculated. This is also essential if solvent molecules occur. Since this is a key issue of modern quantum chemical studies, we are focusing here on the sublinear calculations of selected nuclei in a chemical system. To exploit the sublinear-scaling behavior, the numerically significant elements of the two-electron integrals in the rate-determining steps have to be determined by using integral estimates. Therefore, we extended the QQR-type integral estimates which have been developed for HF and AO-MP2 energies [7, 8], to the calculations of NMR shieldings.

In this work, we first briefly summarize the rate-determining steps of the AO-MP2 NMR shielding calculations and introduce the QQR-type estimates for reducing the computational cost of these steps to sublinear for single nuclei. The sublinear-scaling behavior and the differences of our AO-MP2 NMR shieldings compared to conventional MO-MP2 NMR shieldings are presented for linear alkanes and amylose chains.

2 The AO-MP2 NMR shieldings

NMR shieldings can be obtained by differentiating the energy with respect to the magnetic field and the nuclear magnetic spin moment. Our AO-MP2 NMR shieldings are the second derivative of the AO-MP2 energy expression, which was introduced by Almlöf and Häser [9–11]:

$$E_{\text{AO-MP2}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} E_{JK}^{(\alpha)} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \sum_{\mu\nu\lambda\sigma} \left(\underline{\mu}\bar{\nu} | \underline{\lambda}\bar{\sigma} \right) \left[2 \left(\mu\nu | \lambda\sigma \right) - \left(\mu\sigma | \lambda\nu \right) \right] , \quad (1)$$

where the transformed two-electron integrals are calculated with the local pseudo-densities:

$$\left(\underline{\mu}\bar{\nu} | \underline{\lambda}\bar{\sigma} \right) = \sum_{\mu'} P_{\mu'\mu} \left(\sum_{\nu'} \bar{P}_{\nu'\nu} \left(\sum_{\lambda'} P_{\lambda'\lambda} \left(\sum_{\sigma'} \bar{P}_{\sigma'\sigma} \left(\mu'\nu' | \lambda'\sigma' \right) \right) \right) \right) . \quad (2)$$

As shown in Ref. [12], we formulated the second derivative of this energy expression with respect to the magnetic field and the nuclear magnetic spin moment:

$$\begin{aligned} \sigma_j^{MP2} = & - \sum_{\alpha=1}^{\tau} w_{\alpha} \left\{ \mathcal{I}^{\mathbf{B}\mathbf{m}_j} + 2\text{Tr} \left[\left(\bar{\mathbf{Y}}_1^{\mathbf{m}_j} - \underline{\mathbf{Y}}_1^{\mathbf{m}_j} + \mathbf{G} [\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \underline{\mathbf{Y}}_2^{\mathbf{m}_j}] + \bar{\mathbf{R}}^{\mathbf{m}_j} e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}} \right. \right. \right. \\ & \left. \left. \left. + \bar{\mathbf{R}} (e^{t_{\alpha}\mathbf{P}_{\text{occ}}\mathbf{F}})^{\mathbf{m}_j} - \underline{\mathbf{R}}^{\mathbf{m}_j} e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} - \underline{\mathbf{R}} (e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j} \right) \mathbf{P}_{\text{occ}}^{\mathbf{B}} \right] \right. \\ & \left. + 2\text{Tr} \left[\mathcal{P} \mathbf{P}_{\text{occ}}^{\mathbf{B}\mathbf{m}_j} \right] \right. \\ & \left. + 2\text{Tr} \left[\left(\bar{\mathbf{Y}}_2^{\mathbf{m}_j} + \underline{\mathbf{Y}}_2^{\mathbf{m}_j} \right) \mathbf{F}^{(\mathbf{B})} + \left(\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2 \right) \mathbf{F}^{(\mathbf{B}\mathbf{m}_j)} \right] \right. \\ & \left. + 2\text{Tr} \left[\left(-(\underline{\mathbf{Y}}_1^{\mathbf{m}_j} + \underline{\mathbf{R}}^{\mathbf{m}_j} e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}} + \underline{\mathbf{R}} (e^{-t_{\alpha}\mathbf{P}_{\text{virt}}\mathbf{F}})^{\mathbf{m}_j}) \mathbf{S}^{-1} \mathbf{S}^{\mathbf{B}} \mathbf{S}^{-1} \right) \right] \right\} , \quad (3) \end{aligned}$$

using the following matrices:

$$\begin{aligned}
\bar{\mathbf{R}}_{\mu'\mu} &= \sum_{\nu\lambda\sigma} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) \\
\underline{\mathbf{R}}_{\nu'\nu} &= \sum_{\mu\lambda\sigma} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) \\
\mathbf{R}_{\nu'\nu}^{\mathbf{m}_j} &= \sum_{\mu\lambda\sigma} \left(\sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'\nu'|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) + \sum_{\lambda'} \underline{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\underline{\mu}\nu'|\lambda'\bar{\sigma}) (\mu\nu||\lambda\sigma) \right. \\
&\quad \left. + \sum_{\sigma'} \bar{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\underline{\mu}\nu'|\underline{\lambda}\sigma') (\mu\nu||\lambda\sigma) \right) \\
&= \sum_{\mu\lambda\sigma} \left((\underline{\mu}^{\mathbf{m}_j}\nu'|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) + (\underline{\mu}\nu'|\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}) (\mu\nu||\lambda\sigma) + (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}) (\mu\nu||\lambda\sigma) \right) \\
\bar{\mathbf{R}}_{\mu'\mu}^{\mathbf{m}_j} &= \sum_{\nu\lambda\sigma} \left(\sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\mu'\nu'|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) + \sum_{\lambda'} \underline{P}_{\lambda'\lambda}^{\mathbf{m}_j} (\mu'\bar{\nu}|\lambda'\bar{\sigma}) (\mu\nu||\lambda\sigma) \right. \\
&\quad \left. + \sum_{\sigma'} \bar{P}_{\sigma'\sigma}^{\mathbf{m}_j} (\mu'\bar{\nu}|\underline{\lambda}\sigma') (\mu\nu||\lambda\sigma) \right) \\
&= \sum_{\nu\lambda\sigma} \left((\mu'\bar{\nu}^{\mathbf{m}_j}|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma) + (\mu'\bar{\nu}|\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}) (\mu\nu||\lambda\sigma) + (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}) (\mu\nu||\lambda\sigma) \right) \\
\mathcal{I}^{\mathbf{Bm}_j} &= 4 \sum_{\mu\nu\lambda\sigma} \sum_{\mu'} \underline{P}_{\mu'\mu}^{\mathbf{m}_j} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}} + 4 \sum_{\mu\nu\lambda\sigma} \sum_{\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}} \\
&= 4 \sum_{\mu\nu\lambda\sigma} (\underline{\mu}^{\mathbf{m}_j}\bar{\nu}|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}} + 4 \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}^{\mathbf{m}_j}|\underline{\lambda}\bar{\sigma}) (\mu\nu||\lambda\sigma)^{\mathbf{B}}.
\end{aligned} \tag{4}$$

Here, the second derivative of the density matrix $\mathbf{P}_{\text{occ}}^{\mathbf{Bm}_j}$ appears, which is expensive to calculate. The explicit calculation of this matrix however, can be avoided using our AO-based formulation [12] of the Z-vector method of Handy and Schaefer [13,14]. Since we are aiming for sublinear-scaling calculations of selected nuclei in large molecules, all rate-determining linear-scaling steps must be avoided as well. Therefore, the AO-based Z-vector method is also applied for avoiding the calculation of the linear-scaling first derivative of the density matrix with respect to the magnetic field $\mathbf{P}_{\text{occ}}^{\mathbf{B}}$ in Eq. 3 (see Ref. [12]).

The remaining rate-determining steps for the AO-MP2 NMR shieldings are now the determination of the matrices in Eq. 4. As shown in Ref. [12], the scaling behavior of these steps becomes sublinear for sufficiently large molecules. To exploit this sublinear-scaling behavior, we developed QQR-type integral estimates for AO-MP2 shieldings, which are presented in the following section.

2.1 Integral estimates for AO-MP2 NMR shieldings

The most common integral estimates for two-electron integrals are the classical Schwarz estimates [15]:

$$|(\mu\nu|\lambda\sigma)| \leq \underbrace{|(\mu\nu|\mu\nu)|^{\frac{1}{2}}}_{Q_{\mu\nu}} \underbrace{|(\lambda\sigma|\lambda\sigma)|^{\frac{1}{2}}}_{Q_{\lambda\sigma}}. \quad (5)$$

Since for the AO-MP2 calculations also transformed integrals with pseudo-densities are needed, Häser introduced the so-called pseudo-Schwarz estimates [11]:

$$\begin{aligned} X_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\lambda\nu|\sigma\nu) P_{\lambda\mu} P_{\sigma\mu} \right)^{\frac{1}{2}} = (\underline{\mu}\nu|\underline{\mu}\nu)^{\frac{1}{2}} \\ Y_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\mu\lambda|\mu\sigma) \bar{P}_{\lambda\nu} \bar{P}_{\sigma\nu} \right)^{\frac{1}{2}} = (\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})^{\frac{1}{2}} \\ Z_{\mu\nu} &= \min \left(\sum_{\lambda} X_{\mu\lambda} |\bar{P}_{\lambda\nu}|, \sum_{\lambda} |P_{\mu\lambda}| Y_{\lambda\nu} \right) \geq (\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})^{\frac{1}{2}}. \end{aligned} \quad (6)$$

Similar to these pseudo-Schwarz estimates, we introduce here pseudo-Schwarz matrices for transformed integrals with perturbed pseudo-densities with respect to the nuclear magnetic spin moment:

$$\begin{aligned} X_{\underline{\mu}^{\mathbf{m}_j}\nu} &= \left(\sum_{\lambda\sigma} (\lambda\nu|\sigma\nu) P_{\lambda\mu}^{\mathbf{m}_j} P_{\sigma\mu}^{\mathbf{m}_j} \right)^{\frac{1}{2}} = (\underline{\mu}^{\mathbf{m}_j}\nu|\underline{\mu}^{\mathbf{m}_j}\nu)^{\frac{1}{2}} \\ Y_{\underline{\mu}\bar{\nu}^{\mathbf{m}_j}} &= \left(\sum_{\lambda\sigma} (\mu\lambda|\mu\sigma) \bar{P}_{\lambda\nu}^{\mathbf{m}_j} \bar{P}_{\sigma\nu}^{\mathbf{m}_j} \right)^{\frac{1}{2}} = (\underline{\mu}\bar{\nu}^{\mathbf{m}_j}|\underline{\mu}\bar{\nu}^{\mathbf{m}_j})^{\frac{1}{2}} \\ Z_{\underline{\mu}\bar{\nu}^{\mathbf{m}_j}} &= \left(\sum_{\lambda} X_{\mu\lambda} |\bar{P}_{\lambda\nu}^{\mathbf{m}_j}| \right) \\ Z_{\underline{\mu}^{\mathbf{m}_j}\bar{\nu}} &= \left(\sum_{\lambda} |P_{\mu\lambda}^{\mathbf{m}_j}| Y_{\lambda\nu} \right) \end{aligned} \quad (7)$$

Schwarz estimates neglect the dependence of the integral value on the bra-ket separation and therefore often overestimate the integral value and no linear- or sublinear-scaling for AO-MP2 can be achieved. Therefore, the distance between the bra and ket distributions is included in the recently introduced QQR-type estimates. As shown for AO-MP2 energies [8], the asymptotic decay behavior of the bra-ket distance can be deduced from the multipole expansion:

$$\begin{aligned} (\Omega_A|\Omega_B) &= \frac{q_{00}^A q_{00}^B}{R_{AB}} + \frac{q_{00}^A \left(\sum_{j=-1}^1 T'_{00,1j} q_{1j}^B \right)}{R_{AB}^2} \\ &\quad + \frac{\left(\sum_{i=-1}^1 q_{1i}^A T'_{1i,00} \right) q_{00}^B}{R_{AB}^2} + \mathcal{O}(R_{AB}^{-3}), \end{aligned} \quad (8)$$

Here, the contributions of bra and ket are presented by arbitrary charge distributions Ω_A and Ω_B , respectively. The multipole expansion of a fully transformed two-electron integral with pseudo-densities is different to the expansion of a conventional two-electron integral. For the former case, the monopoles q_{00} in the first three terms of Eq. 8 are all zero due to the orthogonality of the occupied and virtual subspace:

$$q_{00}^{\underline{\mu}\bar{\nu}} = S_{\underline{\mu}\bar{\nu}} = \sum_{\mu'\nu'} P_{\mu'\mu} S_{\mu\nu} \bar{P}_{\nu\nu'} = 0 . \quad (9)$$

To obtain QQR-type estimates for AO-MP2 NMR shieldings, the decay behavior of the single integral products in Eq. 4 must also be deduced using the multipole expansion. For the integral products involving perturbed pseudo-densities in Eq. 4, the QQR-type estimates are:

$$\left(\underline{\mu}^{\mathbf{m}_j\nu'}|\underline{\lambda}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{X_{\underline{\mu}^{\mathbf{m}_j\nu'}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}^{\mathbf{m}_j\nu'}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (10)$$

$$\left(\underline{\mu}\nu'|\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{X_{\underline{\mu}\nu'}Z_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (11)$$

$$\left(\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{X_{\underline{\mu}\nu'}Z_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (12)$$

$$\left(\mu'\bar{\nu}^{\mathbf{m}_j}|\underline{\lambda}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{Y_{\mu'\bar{\nu}^{\mathbf{m}_j}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\mu'\bar{\nu}^{\mathbf{m}_j}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (13)$$

$$\left(\mu'\bar{\nu}|\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{Y_{\mu'\bar{\nu}}Z_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\mu'\bar{\nu}} - \text{ext}_{\underline{\lambda}^{\mathbf{m}_j}\bar{\sigma}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (14)$$

$$\left(\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}\right)\left(\mu\nu|\lambda\sigma\right) \approx \frac{Y_{\mu'\bar{\nu}}Z_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\mu'\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}^{\mathbf{m}_j}})(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (15)$$

$$\left(\underline{\mu}^{\mathbf{m}_j}\bar{\nu}|\underline{\lambda}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right)^{\mathbf{B}} \approx \frac{Z_{\underline{\mu}^{\mathbf{m}_j}\bar{\nu}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}^{\mathbf{m}_j}\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (16)$$

$$\left(\underline{\mu}\bar{\nu}^{\mathbf{m}_j}|\underline{\lambda}\bar{\sigma}\right)\left(\mu\nu|\lambda\sigma\right)^{\mathbf{B}} \approx \frac{Z_{\underline{\mu}\bar{\nu}^{\mathbf{m}_j}}Z_{\underline{\lambda}\bar{\sigma}}Q_{\mu\nu}Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\bar{\nu}^{\mathbf{m}_j}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2(R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \quad (17)$$

The asymptotic decay behavior of the transformed integrals is $1/(R')^2$, if the ket part is fully transformed with unperturbed pseudo-densities. This is due to the fact that the monopoles $q_{00}^{\lambda\bar{\sigma}}$ are zero and the first term in the multipole expansion vanishes. As

a result, the second term with an asymptotic decay of $1/(R')^2$ determines the decay behavior. For the other transformed integral and the untransformed integrals, the decay behavior is $1/R'$, since none of the terms in the multipole expansion vanish. The derivative of the two-electron integrals with respect to the magnetic field are estimated with the estimates for the non-perturbed two-electron integrals [1, 16, 17]. The distance R' between the charge distributions in bra and ket is here obtained by difference of the distance R between the two centers and the extents of the two charge distributions. As an example, for the first integral in Eq. 10, we have:

$$R' = R - \text{ext}_{\underline{\mu}^{\text{mj}}_{\nu'}} - \text{ext}_{\underline{\lambda}\bar{\sigma}} \quad (18)$$

The definitions of the different extents used for these QQR-type estimates are shown in Refs. [7, 8] or in appendix A. It is important to mention that the multipole expansion can only be applied if the contributions in bra and ket are well separated. We therefore use the condition $R' > 1$ to determine the applicability of the QQR-type estimates. If the condition is fulfilled the QQR-type estimates are applied, otherwise, only the Schwarz-type estimates are used.

For the estimates of the two unperturbed R-matrices $\bar{\mathbf{R}}_{\mu'\mu}$ and $\mathbf{R}_{\nu'\nu}$ in Eq. 4, we introduced in Ref. [12] a different screening procedure. The R-matrices have normally a linear-scaling number of elements. However, we can truncate the unperturbed R-matrices to a number of sublinear-scaling elements by employing the local behavior of the nuclear magnetic spin perturbation. For this purpose, we have included in the screening procedure sublinear-scaling matrices perturbed with respect to the nuclear magnetic spin. For the $\bar{\mathbf{R}}_{\mu'\mu}$ matrices, the following four matrices are included in the screening procedure:

$$\begin{aligned} (\underline{\mu}'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\underline{\mu}\nu|\underline{\lambda}\sigma) \approx & \text{MAX} \left[(\mathbf{P}_{\text{occ}}^{\text{mj}})_{\mu, \text{max}}, (\mathbf{P}_{\text{occ}}^{\text{mj}})_{\text{max}, \mu'}, (\mathbf{P}_{\text{occ}} \mathbf{h}^{\text{Bmj}})_{\mu, \text{max}}, (\mathbf{F}^{\text{mj}} \mathbf{P}_{\text{occ}})_{\text{max}, \mu'} \right] \\ & \cdot \frac{Y_{\mu'\bar{\nu}} Z_{\underline{\lambda}\bar{\sigma}} Q_{\underline{\mu}\nu} Q_{\underline{\lambda}\sigma}}{(R - \text{ext}_{\underline{\mu}'\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2 (R - \text{ext}_{\underline{\mu}\nu} - \text{ext}_{\underline{\lambda}\sigma})} , \end{aligned} \quad (19)$$

and for the $\mathbf{R}_{\nu'\nu}$ matrices the analogous four matrices are:

$$\begin{aligned} (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\underline{\mu}\nu|\underline{\lambda}\sigma) \approx & \text{MAX} \left[(\mathbf{P}_{\text{virt}}^{\text{mj}})_{\nu, \text{max}}, (\mathbf{P}_{\text{virt}}^{\text{mj}})_{\text{max}, \nu'}, (\mathbf{P}_{\text{virt}} \mathbf{h}^{\text{Bmj}})_{\nu, \text{max}}, (\mathbf{F}^{\text{mj}} \mathbf{P}_{\text{virt}})_{\text{max}, \nu'} \right] \\ & \cdot \frac{X_{\underline{\mu}\nu'} Z_{\underline{\lambda}\bar{\sigma}} Q_{\underline{\mu}\nu} Q_{\underline{\lambda}\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2 (R - \text{ext}_{\underline{\mu}\nu} - \text{ext}_{\underline{\lambda}\sigma})} . \end{aligned} \quad (20)$$

Here, the QQR-type estimates for the triply-transformed integrals have a decay behavior of $1/(R')^2$, since the the monopoles $q_{00}^{\lambda\bar{\sigma}}$ are again zero.

3 Computational Details

The calculation of AO-MP2 shieldings employing the present screening procedure is implemented in a development version of the program package Q-Chem [18]. To calculate the deviations between the AO-MP2 and the MO-MP2-based shieldings, calculations of the MO-MP2 shieldings were performed with the TURBOMOLE program package [19]. Furthermore, the scaling behavior is determined with respect to the next larger molecule, where the system size is presented by the number of basis functions. For efficiency reasons, the unperturbed and perturbed pseudo-densities are scaled with the coefficients of the Laplace expansion. Moreover, we define three different screening thresholds: ϑ_1 for the estimates of the integral products in Eqs. 10-15, ϑ_2 for the estimates of the the integral products in Eqs. 16 and 17, and ϑ_3 for the R-matrices in Eqs. 19 and 20. The timings for the AO-MP2 shieldings were performed on a single core of an Intel Xeon E5-2620 using 128 GB of RAM.

4 Results

To demonstrate the sublinear-scaling behavior of the rate-determining steps, we have performed calculations of AO-MP2 NMR shieldings for linear alkanes and amylose chains. As shown in Fig. 1, we select several nuclei in the smallest systems of the linear alkanes and the amylose chains. To extract the scaling behavior, we increase the size of the smallest systems systematically, while the shieldings of the same selected nuclei in all systems is calculated. The scaling behavior of the wall time for forming the terms in Eq. 4 is determined and the deviations of the AO-MP2 results are compared to the conventional MO-MP2 NMR shieldings for all systems.

The results for the linear alkanes are shown in the Tables 1-5. The asymptotic sublinear-scaling behavior for the timings of the first carbon atom of the linear alkanes with the STO-3G basis set are shown for different screening thresholds. As the differences between the AO-MP2 and MO-MP2 results are systematically dependent on the size of the screening thresholds the accuracy of the AO-MP2 NMR shieldings can be controlled. Deviations between the AO-MP2 and the MO-MP2 results of up to 0.1, 0.4 and

1.3 ppm are obtained for the $\vartheta_{1/2/3} = (10^{-8}/10^{-10}/10^{-8})$, $\vartheta_{1/2/3} = (10^{-8}/10^{-8}/10^{-6})$, and $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$ thresholds, respectively. The asymptotic sublinear-scaling behavior is also shown for the sixth carbon atom of linear alkanes (see Tab. 4). While the wall times of the sixth carbon atom are larger than those of the first carbon atom, the scaling behaviors of the two is quite similar. For the 6-31G* basis set, the data are presented for the first carbon atom of linear alkanes and the $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$ thresholds, which show a similar behavior as the results for calculations with the basis set STO-3G and the $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$ thresholds.

The asymptotic sublinear-scaling behavior of the selected nucleus in amylose chains (basis set: STO-3G) is shown in the Tables 6 and 7. Here, the deviations between the AO-MP2 and the MO-MP2 results of up to 0.1 and 0.8 ppm are obtained for the $\vartheta_{1/2/3} = (10^{-8}/10^{-8}/10^{-6})$ and $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$ thresholds, respectively.

5 Conclusion

We have introduced QQR-type integral estimates for AO-MP2 NMR shieldings. By applying these estimates, the computational cost of the rate-determining steps of AO-MP2 NMR calculations of a specific nucleus can be reduced to sublinear, as it is shown for linear alkanes and amylose chains. Furthermore, the differences between the AO-MP2 and MO-MP2 NMR shieldings are explored, which are systematically dependent on the size of the screening thresholds.

Acknowledgements

C. O. acknowledges financial support by the 'Deutsche Forschungsgemeinschaft' (DFG) funding proposal Oc35/4-1.

A Definition of the extents

For the AO-MP2 shieldings, we need different extents for fully-transformed and half-transformed bras and kets, which are transformed with unperturbed or perturbed pseudo-densities. The definition of the untransformed extents and of the fully-transformed extents with unperturbed pseudo-densities are shown in Ref. [8], since these extents are also needed for AO-MP2 energies. For the half-transformed bras or kets, the definition of the extents need to be adapted. The extents for half-transformed bras or kets with the unperturbed virtual pseudo-density are calculated by:

$$\text{ext}_{\mu'\bar{\nu}} = \max_{\{\mu', \nu' | c_{\mu'\bar{\nu}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\mu'\bar{\nu}, \mu'\nu'} + c_{\mu'\bar{\nu}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} , \quad (21)$$

Here, $r_{\mu'\bar{\nu}, \mu'\nu'}$ is the distance between the expansion centers of $\mu'\bar{\nu}$ and $\mu'\nu'$, where the same centers are used for untransformed and transformed bras and kets. For these extents, the relative weight factors are defined as:

$$c_{\mu'\bar{\nu}}^{\mu'\nu'} = \frac{|S_{\mu'\nu'} \bar{P}_{\nu'\nu}|}{\sum_{\sigma} |S_{\mu\sigma} \bar{P}_{\sigma\nu}|} . \quad (22)$$

The extents for half-transformed bras or ket with the occupied pseudo-density are obtained by:

$$\text{ext}_{\underline{\mu}\nu'} = \max_{\{\mu', \nu' | c_{\underline{\mu}\nu'}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\underline{\mu}\nu', \mu'\nu'} + c_{\underline{\mu}\nu'}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} , \quad (23)$$

where the related relative weight factors are calculated by:

$$c_{\underline{\mu}\nu'}^{\mu'\nu'} = \frac{|P_{\underline{\mu}\mu'} S_{\mu'\nu'}|}{\sum_{\lambda} |P_{\underline{\mu}\lambda} S_{\lambda\nu}|} . \quad (24)$$

Furthermore, we use extents for fully- and half-transformed extents with pseudo-densities, which are perturbed with respect to the nuclear magnetic spin moment. The analogous extents are obtained by:

$$\begin{aligned} \text{ext}_{\mu'\bar{\nu}^{\text{mj}}} &= \max_{\{\mu', \nu' | c_{\mu'\bar{\nu}^{\text{mj}}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\mu'\bar{\nu}^{\text{mj}}, \mu'\nu'} + c_{\mu'\bar{\nu}^{\text{mj}}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} \\ \text{with } c_{\mu'\bar{\nu}^{\text{mj}}}^{\mu'\nu'} &= \frac{|S_{\mu'\nu'} \bar{P}_{\nu'\nu}^{\text{mj}}|}{\sum_{\sigma} |S_{\mu\sigma} \bar{P}_{\sigma\nu}^{\text{mj}}|} , \end{aligned} \quad (25)$$

$$\begin{aligned} \text{ext}_{\underline{\mu}^{\mathbf{m}_j\nu'}} &= \max_{\{\mu',\nu' | c_{\underline{\mu}^{\mathbf{m}_j\nu'}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\underline{\mu}^{\mathbf{m}_j\nu'},\mu'\nu'} + c_{\underline{\mu}^{\mathbf{m}_j\nu'}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} \\ \text{with } c_{\underline{\mu}^{\mathbf{m}_j\nu'}}^{\mu'\nu'} &= \frac{|P_{\underline{\mu}\mu'}^{\mathbf{m}_j} S_{\mu'\nu'}|}{\sum_{\lambda} |P_{\underline{\mu}\lambda}^{\mathbf{m}_j} S_{\lambda\nu}|}, \end{aligned} \quad (26)$$

$$\begin{aligned} \text{ext}_{\underline{\mu}^{\mathbf{m}_j\bar{\nu}}} &= \max_{\{\mu',\nu' | c_{\underline{\mu}^{\mathbf{m}_j\bar{\nu}}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\underline{\mu}^{\mathbf{m}_j\bar{\nu}},\mu'\nu'} + c_{\underline{\mu}^{\mathbf{m}_j\bar{\nu}}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} \\ \text{with } c_{\underline{\mu}^{\mathbf{m}_j\bar{\nu}}}^{\mu'\nu'} &= \frac{|P_{\underline{\mu}\mu'}^{\mathbf{m}_j} S_{\mu'\nu'} \bar{P}_{\nu'\nu}|}{\sum_{\lambda\sigma} |P_{\underline{\mu}\lambda}^{\mathbf{m}_j} S_{\lambda\sigma} \bar{P}_{\sigma\nu}|}, \end{aligned} \quad (27)$$

$$\begin{aligned} \text{ext}_{\underline{\mu}^{\bar{\nu}\mathbf{m}_j}} &= \max_{\{\mu',\nu' | c_{\underline{\mu}^{\bar{\nu}\mathbf{m}_j}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\underline{\mu}^{\bar{\nu}\mathbf{m}_j},\mu'\nu'} + c_{\underline{\mu}^{\bar{\nu}\mathbf{m}_j}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} \\ \text{with } c_{\underline{\mu}^{\bar{\nu}\mathbf{m}_j}}^{\mu'\nu'} &= \frac{|P_{\underline{\mu}\mu'} S_{\mu'\nu'} \bar{P}_{\nu'\nu}^{\mathbf{m}_j}|}{\sum_{\lambda\sigma} |P_{\underline{\mu}\lambda} S_{\lambda\sigma} \bar{P}_{\sigma\nu}^{\mathbf{m}_j}|}. \end{aligned} \quad (28)$$

References

- [1] J. Kussmann and C. Ochsenfeld, *J. Chem. Phys.*, **2007**, *127*, 054103.
- [2] M. Beer, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.*, **2011**, *134*, 074102.
- [3] J. Gauss, *J. Chem. Phys.*, **1993**, *99*, 3629–3643.
- [4] M. Kollwitz and J. Gauss, *Chem. Phys. Lett.*, **1996**, *260*, 639–646.
- [5] M. Kollwitz, M. Häser, and J. Gauss, *J. Chem. Phys.*, **1998**, *108*, 8295–8301.
- [6] D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, and C. Ochsenfeld, *J. Chem. Theory Comput.*, **2014**, *10*, 572–578.
- [7] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, *J. Chem. Phys.*, **2012**, *136*, 144107.
- [8] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.*, **2013**, *138*, 014101.
- [9] J. Almlöf, *Chem. Phys. Lett.*, **1991**, *181*, 319–320.
- [10] M. Häser and J. Almlöf, *J. Chem. Phys.*, **1992**, *96*, 489.
- [11] M. Häser, *Theoret. Chim. Acta*, **1993**, *87*, 147–173.
- [12] M. Maurer and C. Ochsenfeld, *J. Chem. Phys.*, **2013**, *138*, 174104.
- [13] N. C. Handy and H. F. Schaefer III, *J. Chem. Phys.*, **1984**, *81*, 5031.
- [14] Y. Yamaguchi, J. D. Goddard, Y. Osamura, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press Inc, 1994.
- [15] M. Häser and R. Ahlrichs, *J. Comput. Chem.*, **1989**, *10*, 104.
- [16] H. Horn, H. Weiß, M. Häser, M. Ehrig, and R. Ahlrichs, *J. Comput. Chem.*, **1991**, *12*, 1058–1064.
- [17] C. Ochsenfeld, *Chem. Phys. Lett.*, **2000**, *327*, 216–223.

- [18] Development version of Q-CHEM, www.q-chem.com.
- [19] Turbomole Version 6.3, a quantum chemical program package written by R. Ahlrichs, M. K. Armbruster, R. A. Bachorz, M. Bär, H. P. Baron, R. Bauernschmitt, F. A. Bischoff, S. Bäcker, N. Crawford, P. Deglmann, F. Della Sala, M. Diedenhofen, M. Ehrig, K. Eichkorn, S. Elliott, F. Furche, A. Glöb, F. Haase, M. Häser, C. Hättig, A. Hellweg, S. Höfener, H. Horn, C. Huber, U. Huniar, M. Kattannek, W. Klopper, A. Köhn, C. Kölmel, M. Kollwitz, K. May, P. Nava, C. Ochsenfeld, H. Ohm, M. Pabst, H. Patzelt, D. Rappoport, O. Rubner, A. Schäfer, U. Schneider, M. Sierka, D. P. Tew, O. Treutler, B. Unterreiner, M. von Arnim, F. Weigend, P. Weis, H. Weiss, N. Winter with contributions from M. Dolg, J. Gauss, C. van Wüllen, S. Brode and H. Schiffer. For the current version, see <http://www.turbomole.com>.

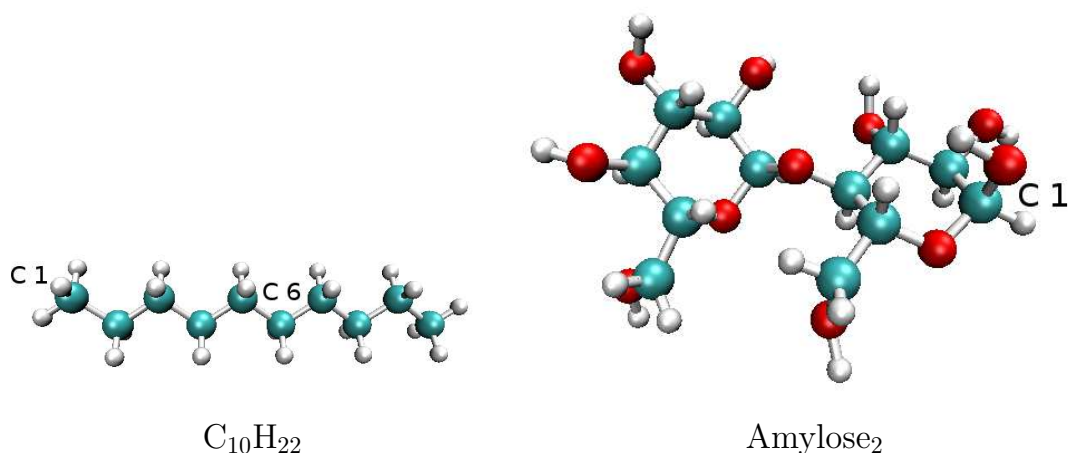


Figure 1: The molecules $\text{C}_{10}\text{H}_{22}$ and amylose_2 are shown, where the selected nuclei for testing the sublinear-scaling AO-MP2 NMR shieldings are labeled with C 1 and C 6.

Table 1: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the first carbon atom (see Fig. 1) of linear alkanes (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-8}/10^{-10}/10^{-8})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
$\text{C}_{10}\text{H}_{22}$ (C 1)	2744	-1.59	-1.59	0.00
$\text{C}_{20}\text{H}_{42}$ (C 1)	11046 (2.1)	-1.59	-1.59	0.00
$\text{C}_{40}\text{H}_{82}$ (C 1)	23048 (1.1)	-1.59	-1.59	0.00
$\text{C}_{80}\text{H}_{162}$ (C 1)	37361 (0.7)	-1.60	-1.59 ^a	0.01
$\text{C}_{200}\text{H}_{402}$ (C 1)	66299 (0.6)	-1.66	-1.59 ^a	0.07

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 2: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the first carbon atom (see Fig. 1) of linear alkanes (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-8}/10^{-8}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
C ₁₀ H ₂₂ (C 1)	2337	-1.59	-1.59	0.00
C ₂₀ H ₄₂ (C 1)	8453 (1.9)	-1.59	-1.59	0.00
C ₄₀ H ₈₂ (C 1)	15314 (0.9)	-1.56	-1.59	-0.03
C ₈₀ H ₁₆₂ (C 1)	21473 (0.5)	-1.65	-1.59 ^a	0.06
C ₂₀₀ H ₄₀₂ (C 1)	31631 (0.4)	-1.94	-1.59 ^a	0.35

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 3: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the first carbon atom (see Fig. 1) of linear alkanes (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
C ₁₀ H ₂₂ (C 1)	953	-1.72	-1.59	0.13
C ₂₀ H ₄₂ (C 1)	1825 (1.0)	-2.02	-1.59	0.43
C ₄₀ H ₈₂ (C 1)	2533 (0.5)	-0.27	-1.59	-1.32
C ₈₀ H ₁₆₂ (C 1)	3342 (0.4)	-0.53	-1.59 ^a	-1.06
C ₂₀₀ H ₄₀₂ (C 1)	5460 (0.5)	-1.96	-1.59 ^a	0.37

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 4: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the sixth carbon atom (see Fig. 1) of linear alkanes (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-8}/10^{-8}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
C ₁₀ H ₂₂ (C 6)	2902	-2.00	-2.00	0.00
C ₂₀ H ₄₂ (C 6)	14294 (2.3)	-2.00	-2.00	0.00
C ₄₀ H ₈₂ (C 6)	25827 (0.9)	-1.97	-2.00	-0.03
C ₈₀ H ₁₆₂ (C 6)	34611 (0.4)	-2.03	-2.00 ^a	0.03
C ₂₀₀ H ₄₀₂ (C 6)	47800 (0.4)	-2.39	-2.00 ^a	0.39

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 5: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the first carbon atom (see Fig. 1) of linear alkanes (basis set: 6-31G*, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
C ₁₀ H ₂₂ (C 1)	114584	6.09	6.03	-0.06
C ₂₀ H ₄₂ (C 1)	340366 (1.6)	6.11	6.03	-0.08
C ₄₀ H ₈₂ (C 1)	589056 (0.8)	4.24	6.03 ^a	1.79

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 6: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the terminal carbon atom (marked with C 1 in Fig. 1) of amylose chains (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-8}/10^{-8}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
Amylose ₂ (C 1)	31618	-3.08	-3.08	0.00
Amylose ₄ (C 1)	106367 (1.8)	-3.08	-3.08	0.00
Amylose ₈ (C 1)	253906 (1.3)	-3.08	-3.08 ^a	0.00
Amylose ₁₆ (C 1)	469275 (0.9)	-2.96	-3.08 ^a	-0.12

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

Table 7: Wall times for computing the terms in Eq. 4 and the scaling behavior (in parentheses) for the terminal carbon atom (marked with C 1 in Fig. 1) of amylose chains (basis set: STO-3G, five Laplace points). Here, the following screening thresholds for the different integral products in Eq. 4 (see computational details) are used: $\vartheta_{1/2/3} = (10^{-6}/10^{-6}/10^{-6})$. Furthermore, the differences (Δ) between the correlation contributions of the AO-MP2 and MO-MP2 based NMR shieldings (in ppm) are shown.

system	time [sec]	AO-MP2 shieldings	MO-MP2 shieldings	Δ
Amylose ₂ (C 1)	9142	-3.20	-3.08	0.12
Amylose ₄ (C 1)	17183 (0.9)	-3.34	-3.08	0.26
Amylose ₈ (C 1)	25418 (0.6)	-2.94	-3.08 ^a	-0.14
Amylose ₁₆ (C 1)	34567 (0.4)	-2.28	-3.08 ^a	-0.80

^a Since the MO-MP2 NMR shielding calculation was not possible, the value was extrapolated instead.

- 5.5** Manuscript V: "QQR-type integral estimates for calculating linear-scaling energy gradients in atomic orbital-based second-order Møller-Plesset perturbation theory",
M. Maurer and C. Ochsenfeld,
(in preparation)

Comment

This paper introduces the theory for the QQR-type estimates for calculating AO-MP2 gradients. Since this work is currently under preparation, only preliminary results, mainly using the small basis set STO-3G, are presented for larger molecules to show the asymptotic scaling behavior. For smaller molecules, results are also presented for the basis set 6-31G* which show a similar scaling behavior to those of STO-3G. Overall, one can expect also a similar scaling behavior for larger molecules, for which calculations are currently ongoing.

**QQR-type integral estimates for calculating
linear-scaling energy gradients in atomic orbital-based
second-order Møller-Plesset perturbation theory**

Marina Maurer and Christian Ochsenfeld^{a)}

*Chair of Theoretical Chemistry, Department of Chemistry,
University of Munich (LMU), Butenandtstr. 7, D-81377 Munich, Germany
and*

*Center for Integrated Protein Science (CIPSM) at the Department of Chemistry,
University of Munich (LMU), Butenandtstr. 5-13, D-81377 Munich, Germany*

QQR-type estimates are introduced for our fully atomic orbital (AO-) based energy gradients in second-order Møller-Plesset perturbation theory (MP2). By applying these estimates the computational cost of the rate-determining steps in AO-MP2 gradients can be reduced to linear, where the accuracy can be fully controlled by the screening threshold. The reliability of our QQR-type estimates for AO-MP2 gradients is demonstrated on linear alkanes, amylose chains, and DNA double-strands.

^{a)}Electronic mail: christian.ochsenfeld@uni-muenchen.de

1 Introduction

Analytic energy gradients are widely used in quantum chemistry to provide equilibrium and transition structures of a molecule. While Hartree-Fock (HF) energy gradients include the electron-electron interaction over a mean field, the missing electron correlation effects can be included, e.g., with the coupled-cluster (CC) or second-order Møller-Plesset perturbation theory (MP2) methods. However, the computational effort of these so-called electron correlation methods is much higher, where even for the cheapest MP2 method the effort increases with $\mathcal{O}(M^5)$ with M being the system size. Therefore, much effort has been made to overcome the high computational effort of the conventional MP2 method. There are several approaches to improve the efficiency of the molecular-orbital (MO-) based MP2 method by combining it with the resolution-of-the-identity (RI) approximation [1–4] or local correlation approximations [5–7].

We exploit here an other approach, which is based on the atomic-orbital (AO-) based MP2 method, originally introduced by Almlöf and Häser [8–10]. In 2008, we introduced the theory of an overall linear-scaling reformulation of the AO-MP2 energy gradients [11] where any transformations between the AO and MO basis are avoided. Within this work, the working equations and first results of a preliminary implementation without integral estimates were presented. Here, we introduce the integral estimates for the AO-MP2 gradients which are based on the recently introduced QQR-type estimates [12, 13]. By using these estimates, Maurer et al. performed calculations of AO-MP2 energies in a linear-scaling fashion for molecules with more than 1000 atoms and 10 000 basis functions [13].

First, we briefly summarize the main equations for the AO-MP2 gradients and introduce a modified version of the QQR-type estimates adapted for gradients. Moreover, we show the scaling behavior of the rate-determining steps for linear alkanes, amylose with different chain lengths, and for DNA double-strands, as obtained using the here introduced QQR-type estimates for the AO-MP2 gradients. Furthermore, the accuracy of our approach is compared with conventional MO-MP2 gradients.

2 Theory

2.1 AO-MP2 energy gradients

As introduced by Almlöf and Häser [8–10], the AO-MP2 energy is calculated by:

$$E_{\text{AO-MP2}} = - \sum_{\alpha=1}^{\tau} w_{\alpha} E_{JK}^{(\alpha)} = - \sum_{\alpha=1}^{\tau} w_{\alpha} \sum_{\mu\nu\lambda\sigma} \left(\underline{\mu}\bar{\nu} | \underline{\lambda}\bar{\sigma} \right) \left[2 \left(\mu\nu | \lambda\sigma \right) - \left(\mu\sigma | \lambda\nu \right) \right] . \quad (1)$$

Thereby, the two-electron integrals are transformed with the local pseudo-densities

$$\left(\underline{\mu}\bar{\nu} | \underline{\lambda}\bar{\sigma} \right) = \sum_{\mu'} \underline{P}_{\mu'\mu} \left(\sum_{\nu'} \bar{P}_{\nu'\nu} \left(\sum_{\lambda'} \underline{P}_{\lambda'\lambda} \left(\sum_{\sigma'} \bar{P}_{\sigma'\sigma} \left(\mu'\nu' | \lambda'\sigma' \right) \right) \right) \right) . \quad (2)$$

To obtain the energy gradients, the energy expression is differentiated with respect to the nuclear coordinates ξ . As shown by Schweizer et al. [11], the first derivative of the AO-MP2 energy can be written as

$$\begin{aligned} E_{\text{AO-MP2}}^{\xi} &= - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2 \mathcal{I}^{\xi} \\ &\quad + 2 \text{Tr} \left[\left(\bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G} [\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] + \bar{\mathbf{R}} e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}} - \underline{\mathbf{R}} e^{-t_{\alpha} \mathbf{P}_{\text{virt}} \mathbf{F}} \right) \mathbf{P}_{\text{occ}}^{\xi} \right] \\ &\quad + 2 \text{Tr} \left[\left(\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2 \right) \mathbf{F}^{(\xi)} \right] \\ &\quad + 2 \text{Tr} \left[- \left(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}} e^{-t_{\alpha} \mathbf{P}_{\text{virt}} \mathbf{F}} \right) \mathbf{S}^{-1} \mathbf{S}^{\xi} \mathbf{S}^{-1} \right] \} \\ &= - \sum_{\alpha=1}^{\tau} w_{\alpha} \{ 2 \mathcal{I}^{\xi} + 2 \text{Tr} \left[\mathcal{P} \mathbf{P}_{\text{occ}}^{\xi} \right] + 2 \text{Tr} \left[\mathcal{F} \mathbf{F}^{(\xi)} \right] + 2 \text{Tr} \left[\mathcal{S} \mathbf{S}^{-1} \mathbf{S}^{\xi} \mathbf{S}^{-1} \right] \} \end{aligned} \quad (3)$$

with

$$\mathcal{P} = \bar{\mathbf{Y}}_1 - \underline{\mathbf{Y}}_1 + \mathbf{G} [\bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2] + \bar{\mathbf{R}} e^{t_{\alpha} \mathbf{P}_{\text{occ}} \mathbf{F}} - \underline{\mathbf{R}} e^{-t_{\alpha} \mathbf{P}_{\text{virt}} \mathbf{F}} \quad (4a)$$

$$\mathcal{F} = \bar{\mathbf{Y}}_2 + \underline{\mathbf{Y}}_2 \quad (4b)$$

$$\mathcal{S} = - \left(\underline{\mathbf{Y}}_1 + \underline{\mathbf{R}} e^{-t_{\alpha} \mathbf{P}_{\text{virt}} \mathbf{F}} \right) . \quad (4c)$$

The expensive determination of the explicit first derivative of the density matrix $\mathbf{P}_{\text{occ}}^{\xi}$ can be avoided by using our AO-based reformulation of the Z-vector method [11], that had been originally formulated by Handy and Schaefer in the MO basis [14, 15].

The time-determining steps in AO-MP2 calculations are the transformations of the two-electron integral with the pseudo-densities in Eq. 2 and the contractions of the

integral products in Eq. 1. For the AO-MP2 gradients, these steps are included in the calculations for the following terms:

$$\begin{aligned}\mathcal{I}^\xi &= \sum_{\mu\nu\lambda\sigma} (\underline{\mu}\bar{\nu}|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma)^\xi \\ \bar{R}_{\mu'\mu} &= \sum_{\nu\lambda\sigma} (\mu'\bar{\nu}|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma) \\ \underline{R}_{\nu'\nu} &= \sum_{\mu\lambda\sigma} (\underline{\mu}\bar{\nu}'|\lambda\bar{\sigma}) (\mu\nu||\lambda\sigma) .\end{aligned}\tag{5}$$

In order to reduce the scaling behavior of these steps to linear, we present here an extension of the recently introduced QQR-type estimates, which were developed by Maurer et al. for HF and AO-MP2 energies [12, 13].

2.2 Integral estimates for AO-MP2 gradients

By using integral estimates one can preselect numerically significant contributions of two-electron integrals. The standard integral estimates in quantum chemistry are the classical Schwarz estimates [16] for the two-electron integrals:

$$|(\mu\nu|\lambda\sigma)| \leq \underbrace{|(\mu\nu|\mu\nu)|^{\frac{1}{2}}}_{Q_{\mu\nu}} \underbrace{|(\lambda\sigma|\lambda\sigma)|^{\frac{1}{2}}}_{Q_{\lambda\sigma}} .\tag{6}$$

Since integral estimates for transformed integrals with pseudo-densities are also needed for AO-MP2 calculations, Häser introduced the pseudo-Schwarz matrices [10]:

$$\begin{aligned}X_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\lambda\nu|\sigma\nu) \underline{P}_{\lambda\mu} \underline{P}_{\sigma\mu} \right)^{\frac{1}{2}} = (\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})^{\frac{1}{2}} \\ Y_{\mu\nu} &= \left(\sum_{\lambda\sigma} (\mu\lambda|\mu\sigma) \bar{P}_{\lambda\nu} \bar{P}_{\sigma\nu} \right)^{\frac{1}{2}} = (\mu\bar{\nu}|\mu\bar{\nu})^{\frac{1}{2}} \\ Z_{\mu\nu} &= \min \left(\sum_{\lambda} X_{\mu\lambda} |\bar{P}_{\lambda\nu}| , \sum_{\lambda} |\underline{P}_{\mu\lambda}| Y_{\lambda\nu} \right) \geq (\underline{\mu}\bar{\nu}|\underline{\mu}\bar{\nu})^{\frac{1}{2}} .\end{aligned}\tag{7}$$

By using these Schwarz and pseudo-Schwarz estimates the scaling behavior of the required terms in Eq. 5 can be reduced for larger molecules to $\mathcal{O}(N^2)$. However, to achieve linear-scaling the QQR-type estimates have to be employed, which include in contrast to conventional Schwarz estimates the dependence of the integral value on the bra-ket separation [12, 13].

As shown for the QQR-type estimates for AO-MP2 energies [13], the actual decay behavior of the bra-ket separation can be determined by analyzing the terms of the

multipole expansion. The first terms of the multipole expansion are

$$\begin{aligned}
 (\Omega_A|\Omega_B) = & \frac{q_{00}^A q_{00}^B}{R_{AB}} + \frac{q_{00}^A \left(\sum_{j=-1}^1 T'_{00,1j} q_{1j}^B \right)}{R_{AB}^2} \\
 & + \frac{\left(\sum_{i=-1}^1 q_{1i}^A T'_{1i,00} \right) q_{00}^B}{R_{AB}^2} + \mathcal{O}(R_{AB}^{-3}) ,
 \end{aligned} \tag{8}$$

where the basis functions in bra and ket are presented by a charge distribution Ω_A and Ω_B . The first three terms include the monopoles q_{00} . If we are expanding a transformed two-electron integral with pseudo-densities, as shown in Eq. 2, these monopoles are zero, since the occupied and virtual subspace are orthogonal to each other:

$$q_{00}^{\mu\bar{\nu}} = S_{\underline{\mu}\bar{\nu}} = \sum_{\mu'\nu'} P_{\mu'\mu} S_{\mu\nu} \bar{P}_{\nu\nu'} = 0 . \tag{9}$$

In the following, we analyze the integral products for the AO-MP2 gradients in Eq. 5 with the multipole expansion to determine the actual decay behavior of the integral product and to obtain thereby the QQR-type estimates for AO-MP2 gradients.

For the \mathcal{I}^ξ term in Eq. 5 we determine the following QQR-type estimate:

$$\left(\underline{\mu}\bar{\nu} | \underline{\lambda}\bar{\sigma} \right) \left(\mu\nu | \lambda\sigma \right)^\xi \approx \frac{Z_{\underline{\mu}\bar{\nu}} Z_{\underline{\lambda}\bar{\sigma}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^3 (R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \tag{10}$$

Here, for the fully-transformed integral, an asymptotic decay behavior of $1/(R')^3$ is observed, because the monopoles of the multipole expansion $q_{00}^{\mu\bar{\nu}}$ and $q_{00}^{\lambda\bar{\sigma}}$ are zero. Consequently, the first three terms of the multipole expansion vanish and the fourth term controls the decay behavior. The derivative of the two-electron integrals can be estimated using the estimates of the nonperturbed integral [17, 18], for which a decay behavior of $1/R'$ results, since none of the terms vanish.

It has to be noted that the multipole expansion can only be applied if the charge distributions in bra and ket in the expanded integral are well-separated. Therefore, the QQR-type estimates are only used in this case, otherwise the common Schwarz-type estimates are applied. To determine, if the contributions in bra and ket are well-separated, we use the following condition: The difference between the distance R of bra and ket and the extents (ext) of the two contributions in bra and ket has to be greater than one. For example, the following condition must be fulfilled to apply the

QQR-type estimates for the fully transformed integrals:

$$R'_{\underline{\mu}\bar{\nu},\underline{\lambda}\bar{\sigma}} = R - \text{ext}_{\underline{\mu}\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}} > 1 . \quad (11)$$

The extents for the untransformed integrals are calculated, as shown in Appendix B of Ref. [12]. The extents for a fully transformed bra or ket are also used for AO-MP2 energies and calculated by

$$\text{ext}_{\underline{\mu}\bar{\nu}} = \max_{\{\mu',\nu' | c_{\underline{\mu}\bar{\nu}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\underline{\mu}\bar{\nu},\mu'\nu'} + c_{\underline{\mu}\bar{\nu}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} , \quad (12)$$

Here, $r_{\underline{\mu}\bar{\nu},\mu'\nu'}$ is the distance between the expansion centers of $\underline{\mu}\bar{\nu}$ and $\mu'\nu'$. The relative weight factors $c_{\underline{\mu}\bar{\nu}}^{\mu'\nu'}$ are obtained by:

$$c_{\underline{\mu}\bar{\nu}}^{\mu'\nu'} = \frac{|P_{\mu\mu'} S_{\mu'\nu'} \bar{P}_{\nu'\nu}|}{\sum_{\lambda\sigma} |P_{\mu\lambda} S_{\lambda\sigma} \bar{P}_{\sigma\nu}|} . \quad (13)$$

Furthermore, we determine for the integral products for the $\bar{R}_{\mu'\mu}$ and $\underline{R}_{\nu'\nu}$ matrices of Eq. 5 the following QQR-type estimates:

$$\begin{aligned} (\mu'\bar{\nu}|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) &\approx \frac{Y_{\mu'\bar{\nu}} Z_{\underline{\lambda}\bar{\sigma}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\mu'\bar{\nu}} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2 (R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \\ (\underline{\mu}\nu'|\underline{\lambda}\bar{\sigma})(\mu\nu|\lambda\sigma) &\approx \frac{X_{\underline{\mu}\nu'} Z_{\underline{\lambda}\bar{\sigma}} Q_{\mu\nu} Q_{\lambda\sigma}}{(R - \text{ext}_{\underline{\mu}\nu'} - \text{ext}_{\underline{\lambda}\bar{\sigma}})^2 (R - \text{ext}_{\mu\nu} - \text{ext}_{\lambda\sigma})} \end{aligned} \quad (14)$$

The triply-transformed integrals have a decay behavior of $1/(R')^2$, because only the monopoles $q_{00}^{\lambda\bar{\sigma}}$ of the multipole expansion are zero. As a result, the first term is zero and the decay behavior is determined by the second term which has an asymptotic decay behavior of $1/(R')^2$. We therefore need to adapt our extents for the two different half-transformed bra parts. The required extents for the half-transformed bra with the virtual pseudo-density is obtained by:

$$\text{ext}_{\mu'\bar{\nu}} = \max_{\{\mu',\nu' | c_{\mu'\bar{\nu}}^{\mu'\nu'} > \vartheta_t\}} \left\{ r_{\mu'\bar{\nu},\mu'\nu'} + c_{\mu'\bar{\nu}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\} , \quad (15)$$

with the adapted relative weight factors:

$$c_{\mu'\bar{\nu}}^{\mu'\nu'} = \frac{|S_{\mu'\nu'} \bar{P}_{\nu'\nu}|}{\sum_{\sigma} |S_{\mu\sigma} \bar{P}_{\sigma\nu}|} . \quad (16)$$

The other half-transformed extents, involving the occupied pseudo-density are determined by

$$\text{ext}_{\underline{\mu\nu'}} = \max_{\{\mu',\nu' | c_{\underline{\mu\nu'}}^{\mu'\nu'} > \theta_t\}} \left\{ r_{\underline{\mu\nu'},\mu'\nu'} + c_{\underline{\mu\nu'}}^{\mu'\nu'} \text{ext}_{\mu'\nu'} \right\}, \quad (17)$$

where the related relative weight factors are defined as:

$$c_{\underline{\mu\nu'}}^{\mu'\nu'} = \frac{|P_{\underline{\mu\mu'}} S_{\mu'\nu'}|}{\sum_{\lambda} |P_{\underline{\mu\lambda}} S_{\lambda\nu}|}. \quad (18)$$

3 Computational Details

The AO-MP2 gradients using the QQR-type estimates are implemented in a development version of the program package Q-Chem [19]. MO-MP2 gradient calculations are also performed with the program package Q-Chem [19]. Here, the scaling behavior is obtained with respect to the next larger molecule, where the system size is represented by the number of basis functions. For efficiency reasons, the pseudo-densities are scaled with the coefficients of the Laplace expansion. Furthermore, the timings for the AO-MP2 and MO-MP2 gradient calculations were performed on a single core of an Intel Xeon E5-2620 using 128 GB of RAM.

4 Results

To analyze the performance of our AO-MP2 gradients, the scaling behavior is determined for the number of two-electron integrals and the wall time. First, we count the number of elements of the integral products in Eq. 5, which are preselected with our QQR-type estimates. Furthermore, we determine the total wall time for (i) applying the QQR-type estimates to obtain the list of significant integrals and, (ii) for the subsequent calculations of these integrals to form the terms in Eq. 5. To demonstrate the accuracy of the AO-MP2 gradients, the root mean square deviations (RMSD) with respect to conventional MO-MP2 gradients are calculated. Moreover, the wall times for the whole AO-MP2 and MO-MP2 gradient calculations are compared. These data are obtained for linear alkanes, amylose chains, and DNA double-strands.

For the linear alkanes, the results for the STO-3G and 6-31G* basis sets are presented in the Tables 1-8. The number of significant elements for the basis set STO-3G shows a linear-scaling behavior for larger molecules and the RMSD values are under

3 $\mu\text{hartree}/\text{bohr}$ and under 20 $\mu\text{hartree}/\text{bohr}$ for the threshold of 10^{-8} and for the threshold of 10^{-6} , respectively. The timings for the calculations of the terms in Eq. 5 and for the threshold of 10^{-6} show a linear-scaling behavior for larger molecules and a subquadratic-scaling behavior for the threshold of 10^{-8} . The comparison of the total wall times of AO-MP2 and MO-MP2 gradient calculations show a small speedup for the $\text{C}_{80}\text{H}_{162}$ molecule and the threshold of 10^{-6} , which will increase for larger linear alkanes. As for the 6-31G* basis set (Tables 5-8) the scaling behavior is similar to the results with the STO-3G basis set and the RMSD values are under 360 $\mu\text{hartree}/\text{bohr}$ for the shown thresholds. Furthermore, the speedups with respect to the MO-MP2 gradients are also similar for the two basis sets, as shown in Tabs. 3 and 7.

The results for the amylose systems and the DNA double-strands are shown in Tables 9-12 (basis set: STO-3G). The number of significant elements show a linear-scaling behavior for larger molecules and the timings show a subquadratic-scaling behavior. For these systems, we also compared the total wall times of the whole calculation of the AO-MP2 gradient with those of the MO-MP2 gradient (see Tab. 11). In Figs. 1 and 2, the total wall times for calculating the AO-MP2 and the MO-MP2 gradients are plotted against the number of basis functions for the amylose chains and the DNA double-strands, respectively. For the amylose chains, the timings of the AO-MP2 gradients show a crossover with conventional MO-MP2 gradients at system sizes between four and eight glucose units. For a amylose chain with sixteen glucose units the speedup with our AO-MP2 method is roughly a factor of four in comparison to the estimated time for the MO-MP2 gradient calculation. Furthermore, the AO-MP2 gradient calculation of the DNA double-strand with four base pairs is faster by a factor of roughly 18 in comparison to the conventional MO-MP2 calculation.

5 Conclusion

QQR-type estimates for AO-MP2 gradients are introduced. By using these estimates, the number of the integral products needed in the rate-determining steps can be reduced to linear, whereas the differences between the AO-MP2 results and the MO-MP2 results can be fully controlled by a screening threshold. The method was tested on linear alkanes, amylose chains and DNA double-strands. For a DNA double-strand with four base pairs, it is shown, that the speedup of the total wall time for our AO-MP2

gradient calculations (STO-3G basis set) is roughly a factor of 18 in comparison to conventional MO-MP2 gradients calculations, while the RMSD with respect to these MO-MP2 gradients is only 133 μ hartree/bohr.

Acknowledgements

C. O. acknowledges financial support by the 'Deutsche Forschungsgemeinschaft' (DFG) funding proposal Oc35/4-1.

References

- [1] M. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.*, **1993**, *208*, 359–363.
- [2] F. Weigend and M. Häser, *Theor. Chem. Acc.*, **1997**, *97*, 331–340.
- [3] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.*, **1998**, *294*, 143–152.
- [4] D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.*, **1996**, *250*, 477–484.
- [5] S. Saebø, J. Baker, K. Wolinski, and P. Pulay, *J. Chem. Phys.*, **2004**, *120*, 11423–11431.
- [6] A. El Azhary, G. Rauhut, P. Pulay, and H.-J. Werner, *J. Chem. Phys.*, **1998**, *108*, 5185–5193.
- [7] M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby, *J. Chem. Phys.*, **2004**, *121*, 737–750.
- [8] J. Almlöf, *Chem. Phys. Lett.*, **1991**, *181*, 319–320.
- [9] M. Häser and J. Almlöf, *J. Chem. Phys.*, **1992**, *96*, 489.
- [10] M. Häser, *Theoret. Chim. Acta*, **1993**, *87*, 147–173.
- [11] S. Schweizer, B. Doser, and C. Ochsenfeld, *J. Chem. Phys.*, **2008**, *128*, 154101.
- [12] S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld, *J. Chem. Phys.*, **2012**, *136*, 144107.
- [13] S. A. Maurer, D. S. Lambrecht, J. Kussmann, and C. Ochsenfeld, *J. Chem. Phys.*, **2013**, *138*, 014101.
- [14] N. C. Handy and H. F. Schaefer III, *J. Chem. Phys.*, **1984**, *81*, 5031.
- [15] Y. Yamaguchi, J. D. Goddard, Y. Osamura, and H. F. Schaefer III, *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*, Oxford University Press Inc, 1994.

- [16] M. Häser and R. Ahlrichs, *J. Comput. Chem.*, **1989**, *10*, 104.
- [17] H. Horn, H. Weiß, M. Häser, M. Ehrig, and R. Ahlrichs, *J. Comput. Chem.*, **1991**, *12*, 1058–1064.
- [18] C. Ochsenfeld, *Chem. Phys. Lett.*, **2000**, *327*, 216–223.
- [19] Development version of Q-CHEM, www.q-chem.com.

Table 1: Number of elements and scaling behavior (in parentheses) of the integral products in Eq. 5, which are determined by applying the QQR-type estimates for the AO-MP2 gradients. The results are shown for linear alkane systems (basis set STO-3G, first Laplace point) and different screening thresholds ϑ .

$\vartheta = 10^{-8}$	\mathcal{I}^ξ		$\overline{R}_{\mu'\mu}$				$\underline{R}_{\nu'\nu}$			
System	$(\underline{\mu}\bar{\nu} \underline{\lambda}\bar{\sigma})(\mu\nu \lambda\sigma)^\xi$		$(\mu'\bar{\nu} \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$		$(\underline{\mu}\nu' \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$	
C ₁₀ H ₂₂	7858744		16782970		11273608		16639816		11492686	
C ₂₀ H ₄₂	22214760	(1.5)	80126830	(2.3)	40307944	(1.9)	78881816	(2.3)	41138796	(1.9)
C ₄₀ H ₈₂	51026240	(1.2)	225672488	(1.5)	105069580	(1.4)	220811634	(1.5)	106876234	(1.4)
C ₈₀ H ₁₆₂	108650252	(1.1)	523147266	(1.2)	238450950	(1.2)	509673842	(1.2)	241878804	(1.2)

$\vartheta = 10^{-6}$	\mathcal{I}^ξ		$\overline{R}_{\mu'\mu}$				$\underline{R}_{\nu'\nu}$			
System	$(\underline{\mu}\bar{\nu} \underline{\lambda}\bar{\sigma})(\mu\nu \lambda\sigma)^\xi$		$(\mu'\bar{\nu} \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$		$(\underline{\mu}\nu' \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$	
C ₁₀ H ₂₂	3869264		9433746		6365392		9149476		6538496	
C ₂₀ H ₄₂	9594574	(1.3)	29583998	(1.7)	17481208	(1.5)	28510360	(1.7)	17974984	(1.5)
C ₄₀ H ₈₂	21043370	(1.1)	70127202	(1.3)	39740734	(1.2)	67474918	(1.3)	40882842	(1.2)
C ₈₀ H ₁₆₂	43939602	(1.1)	151217805	(1.1)	84260134	(1.1)	145402590	(1.1)	86700656	(1.1)

Table 2: Total wall times for (i) applying the QQR-type estimates to obtain the list of significant integrals and, (ii) for the subsequent calculations of these integrals to form the terms in Eq. 5 are shown for linear alkanes with different screening thresholds (basis set: STO-3G, five Laplace points). The corresponding scaling behavior is presented in parentheses.

System	time [sec]			
	$\vartheta = 10^{-8}$		$\vartheta = 10^{-6}$	
C ₁₀ H ₂₂	406		193	
C ₂₀ H ₄₂	2018	(2.4)	693	(1.9)
C ₄₀ H ₈₂	6799	(1.8)	1931	(1.5)
C ₈₀ H ₁₆₂	20240	(1.6)	5172	(1.4)

Table 3: Total wall times for AO-MP2 and MO-MP2 gradients calculations are shown for linear alkanes with different screening thresholds (basis set: STO-3G, five Laplace points). The corresponding scaling behavior is presented in parentheses. Furthermore, the MO-MP2/ AO-MP2 total wall time ratio is shown.

System	AO-MP2 gradients						MO-MP2 gradients	
	total time [sec]			ratio			total time [sec]	
	$\vartheta = 10^{-8}$			$\vartheta = 10^{-6}$				
C ₁₀ H ₂₂	494		0.0	283		0.1	15	
C ₂₀ H ₄₂	2456	(2.4)	0.1	1125	(2.0)	0.1	140	(3.3)
C ₄₀ H ₈₂	8770	(1.9)	0.2	3884	(1.8)	0.4	1640	(3.6)
C ₈₀ H ₁₆₂	29394	(1.8)	0.5	14226	(1.9)	1.1	15900	(3.3)

Table 4: Root mean square deviation (RMSD) in [μ hartree/bohr] of AO-MP2 based gradients (basis set STO-3G, five Laplace points) with respect to conventional MO-MP2 gradients for linear alkane systems and for different screening threshold ϑ .

System	RMSD	
	$\vartheta = 10^{-8}$	$\vartheta = 10^{-6}$
C ₁₀ H ₂₂	2.815	12.787
C ₂₀ H ₄₂	2.449	18.347
C ₄₀ H ₈₂	2.314	19.318
C ₈₀ H ₁₆₂	2.313	19.965

Table 5: Number of elements and scaling behavior (in parentheses) of the integral products in Eq. 5, which are determined by applying the QQR-type estimates for the AO-MP2 gradients. The results are shown for linear alkane systems (basis set 6-31G*, first Laplace point). A screening threshold of 10^{-8} is applied for the \mathcal{I}^ξ -term and a threshold of 10^{-6} is used for the R-matrices.

System	\mathcal{I}^ξ		$\bar{R}_{\mu'\mu}$				$\underline{R}_{\nu'\nu}$			
	$(\underline{\mu}\bar{\nu} \underline{\lambda}\bar{\sigma})(\mu\nu \lambda\sigma)^\xi$		$(\underline{\mu}'\bar{\nu} \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$		$(\underline{\mu}\nu' \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$	
C ₁₀ H ₂₂	362530636		600619910		376740738		421719510		278755062	
C ₂₀ H ₄₂	1099802602	1.6	2294835191	2.0	1216539298	1.7	1330833292	1.7	809612666	1.6
C ₄₀ H ₈₂	2598949690	1.3	5852934173	1.4	2943088050	1.3	3168590606	1.3	1878470936	1.2

Table 6: Total wall times for (i) applying the QQR-type estimates to obtain the list of significant integrals and, (ii) for the subsequent calculations of these integrals to form the terms in Eq. 5 are shown for linear alkanes (basis set: 6-31G*, five Laplace points). The corresponding scaling behavior is presented in parentheses. A screening threshold of 10^{-8} is applied for the \mathcal{I}^ξ -term and a threshold of 10^{-6} is used for the R-matrices.

System	time [sec]
C ₁₀ H ₂₂	15490
C ₂₀ H ₄₂	81244 (2.4)
C ₄₀ H ₈₂	296125 (1.9)

Table 7: Total wall times for AO-MP2 and MO-MP2 gradients calculations are shown for linear alkanes (basis set: 6-31G*, five Laplace points). The corresponding scaling behavior is presented in parentheses. A screening threshold of 10^{-8} is applied for the \mathcal{I}^ξ -term and a threshold of 10^{-6} is used for the R-matrices. Furthermore, the MO-MP2/ AO-MP2 total wall time ratio is presented.

System	AO-MP2 gradients		MO-MP2 gradients	
	total time [sec]	ratio	total time [sec]	
C ₁₀ H ₂₂	17028	0.0	269	
C ₂₀ H ₄₂	90234 (2.4)	0.0	3050 (3.6)	
C ₄₀ H ₈₂	329923 (1.9)	0.5	157000 (5.7)	

Table 8: Root mean square deviation (RMSD) in [μ hartree/bohr] of AO-MP2 based gradients (basis set 6-31G*, five Laplace points) with respect to conventional MO-MP2 gradients for linear alkane systems. A screening threshold of 10^{-8} is applied for the \mathcal{I}^ξ -term and a threshold of 10^{-6} is used for the R-matrices.

System	RMSD
C ₁₀ H ₂₂	103.676
C ₂₀ H ₄₂	219.838
C ₄₀ H ₈₂	354.541

Table 9: Number of elements and scaling behavior (in parentheses) of the integral products in Eq. 5, which are determined by applying the QQR-type estimates for the AO-MP2 gradients. The results are shown for amylose systems and for DNA double-strands (basis set: STO-3G, first Laplace point, screening thresholds of 10^{-6}).

System	\mathcal{I}^ξ		$\overline{R}_{\mu'\mu}$				$\underline{R}_{\nu'\nu}$			
	$(\underline{\mu}\bar{\nu} \underline{\lambda}\bar{\sigma})(\mu\nu \lambda\sigma)^\xi$		$(\underline{\mu}'\bar{\nu} \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$		$(\underline{\mu}\nu' \underline{\lambda}\bar{\sigma})$		$(\mu\nu \lambda\sigma)$	
Amylose ₂	19628525		62156838		37958349		60772472		40465457	
Amylose ₄	47000213	(1.3)	169358374	(1.5)	95880066	(1.4)	165204419	(1.5)	101805766	(1.4)
Amylose ₈	101591565	(1.1)	391080633	(1.2)	213674274	(1.2)	381818508	(1.2)	227954614	(1.2)
Amylose ₁₆	210767349	(1.1)	834260520	(1.1)	449666996	(1.1)	815270137	(1.1)	480834821	(1.1)
DNA ₁	32132916		104207610		57289793		101889169		60247352	
DNA ₂	106034336	(1.5)	450844826	(1.9)	224884751	(1.8)	436438406	(1.9)	235596298	(1.8)
DNA ₄	274325292	(1.3)	1293290776	(1.4)	605503349	(1.4)	1245094047	(1.4)	633414294	(1.3)

Table 10: Total wall times for (i) applying the QQR-type estimates to obtain the list of significant integrals and, (ii) for the subsequent calculations of these integrals to form the terms in Eq. 5 are shown for amylose systems and for DNA double-strands (basis set: STO-3G, five Laplace points, screening thresholds of 10^{-6}). The corresponding scaling behavior is presented in parentheses.

System	time [sec]
Amylose ₂	1371
Amylose ₄	4615 (1.8)
Amylose ₈	13292 (1.6)
Amylose ₁₆	38920 (1.6)
DNA ₁	2227
DNA ₂	13340 (2.3)
DNA ₄	55156 (1.9)

Table 11: Total wall times for AO-MP2 and MO-MP2 gradients calculations are shown for amylose systems and for DNA double-strands (basis set: STO-3G, five Laplace points, screening thresholds of 10^{-6}). The corresponding scaling behavior is presented in parentheses. Furthermore, the MO-MP2/ AO-MP2 total wall time ratio is shown.

System	AO-MP2 gradients			MO-MP2 gradients	
	total time [sec]		ratio	total time [sec]	
Amylose ₂	2208		0.1	215	
Amylose ₄	8456	(2.0)	0.2	1890	(3.3)
Amylose ₈	31238	(1.9)	1.0	31300	(4.1)
Amylose ₁₆	125460	(2.0)	4.2	532522 ^a	(4.1) ^a
DNA ₁	4395		0.1	541	
DNA ₂	29751	(2.5)	0.7	20700	(4.7)
DNA ₄	146008	(2.2)	18.3	2670000	(6.6)

^a The MO-MP2 value for Amylose₁₆ is extrapolated with the scaling behavior of the two previous points.

Table 12: Root mean square deviation (RMSD) in [μ hartree/bohr] of AO-MP2 based gradients (basis set STO-3G, five Laplace points, screening threshold of 10^{-6}) with respect to conventional MO-MP2 gradients for amylose systems and for DNA double-strands.

System	RMSD
Amylose ₂	41.160
Amylose ₄	42.879
Amylose ₈	68.499
Amylose ₁₆	— ^a
DNA ₁	60.546
DNA ₂	95.404
DNA ₄	133.367

^a MO-MP2 gradients of Amylose₁₆ can not be calculated.

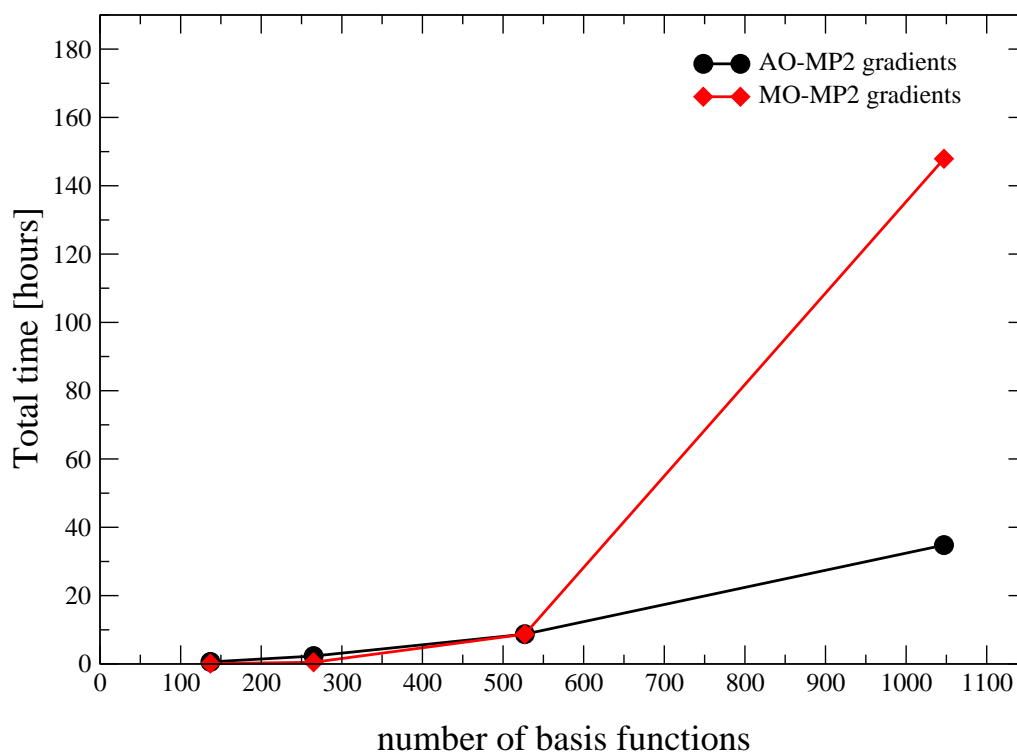


Figure 1: Total wall times of the whole AO-MP2 and MO-MP2 gradient calculations for Amylose₂, Amylose₄, Amylose₈ and Amylose₁₆ (basis set: STO-3G, five Laplace points, screening thresholds of 10^{-6}). The MO-MP2 value for Amylose₁₆ is extrapolated with the scaling behavior of the two previous points.

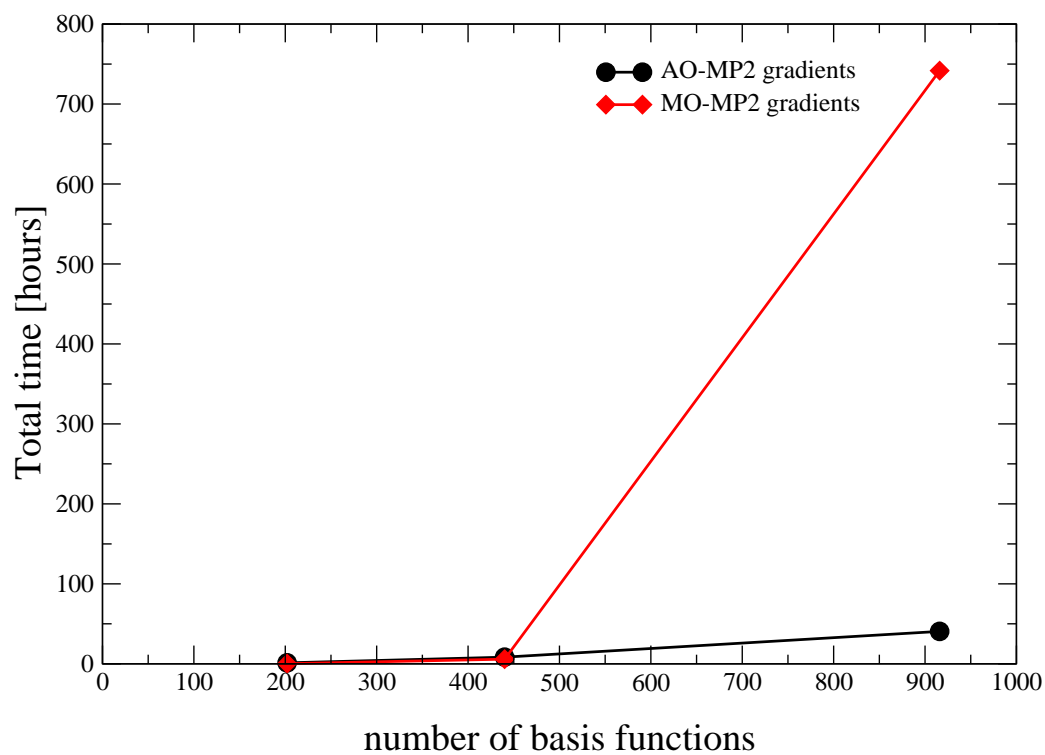


Figure 2: Total wall times of the whole AO-MP2 and MO-MP2 gradient calculations for DNA₁, DNA₂ and DNA₄ (basis set: STO-3G, five Laplace points, screening thresholds of 10^{-6}).

Curriculum vitae

Personal Details

Date of birth	April 28, 1985
Place of birth	Karlsruhe
Nationality	German

Education

since 03/2010	PhD student at the University of Munich (LMU) in the group of Prof. Dr. C. Ochsenfeld
09/2009–02/2010	PhD student at the University of Tübingen in the group of Prof. Dr. C. Ochsenfeld
08/2009	Diploma in chemistry, University of Tübingen (grade: very good) Diploma Thesis: "Entwicklungen zur effizienten Implementierung von AO-MP2-Energiegradienten" Supervisor: Prof. Dr. C. Ochsenfeld
10/2004–08/2009	Studies in chemistry at the University of Tübingen

Awards

06/2013	Poster prize, 7th Molecular Quantum Mechanics conference, Lugano, Switzerland. Poster title: A linear- and sublinear-scaling AO-MP2 reformulation for calculating NMR shieldings
---------	---